

Human Action Recognition in Sports:

A Review of Approaches, Challenges, and Datasets

University of Jordan

OVERVIEW

The paper explores Human Action Recognition (HAR) and its application in sports, focusing on recent progress in computer vision. It emphasizes how HAR can enhance game strategies, player performance, injury prevention, and officiating.

The review discusses the techniques used in HAR, such as feature extraction and action classification, with sports examples like basketball, football, and tennis. It covers the use of Convolutional Neural Networks (CNNs) for image analysis and classifiers like Support Vector Machines (SVM) for action recognition. The paper also highlights challenges in HAR, such as dealing with occlusions and lighting issues that need to be resolved by researchers.

Furthermore, it underscores the significance of large datasets in understanding athletic movements. Overall, the paper aims to provide a comprehensive overview of HAR's role in sports, including the methods used and the challenges faced.





Dr. Tamam Alsarhan
Artificial Intelligence department
King Abdullah || School of Information Technology University of Jordan

Layan Balbisi	0215423
Leen Samman	0219463
Zaina Abunasser	0218080

Table of Contents

1.Introduction.....	4
2.Human Action Recognition	5
2.1. Data Modalities	7
2.2. History of Human Action Recognition	9
2.3. Data collection for HAR tasks	10
2.4. Core Components of HAR	11
2.5. HAR Current Trends.....	12
2.6. Challenges of HAR.....	12
2.7. Future Work	13
3. HAR in sports.....	14
3.1. Applications of HAR in sports.....	15
3.2. Sports	16
3.2.1 Basketball.....	16
A. Approaches.....	16
B. Datasets.....	17
C. Challenges	19
3.2.2. Football	19
A. Approaches.....	20
B. Datasets.....	21
C. Challenges	21
3.2.3. Tennis	21
A. Approaches.....	22
B. Datasets.....	22
C. Challenges	23
4. Conclusion and Discussion	24
References	26

1.Introduction

In the light of the great advancements in computer vision and the increase in visual content, there has been notable interest in utilizing Human Action Recognition (HAR) in several fields to work on their development. One of these fields is sports. Researches focusing on the employment of Human Action Recognition in sports has been growing (see figure1). Various applications have appeared from these researches, showing huge improvements in sports, such as improving game strategies, enhancing player performance, prevent players injuries, and even in the field of officiating and collecting statistics [1]

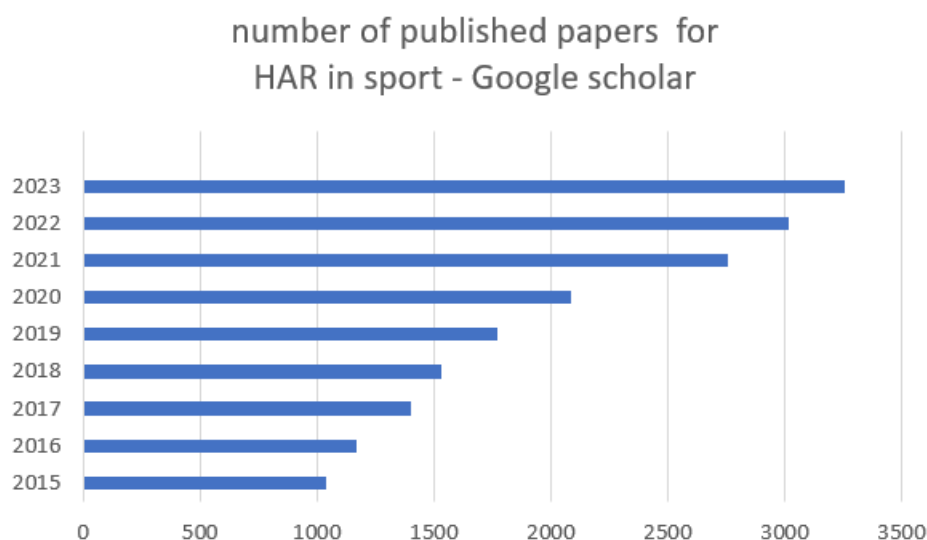


Figure 1 Number of published papers -HAR in sports

This review paper aims to present a clear understanding of how Human Action Recognition (HAR) is used in sports today. The study of movements in sports using several methods and tools will be reviewed, highlighting approaches and challenges experienced. Furthermore, the application of HAR in specific sports such as football, basketball, and tennis to enhance understanding of athletes' movement and performance will be analyzed.

Besides, the data used by researchers to analyze these movements will be discussed. Large, detailed datasets play an essential role in the research as they have an effect in understanding athletic movements and performance.

2.Human Action Recognition

Human action recognition (HAR) is a term in which we use data from videos, sensors, audio, and other various data modalities, to identify and categorize specific actions and behaviors done by individuals, ranging from simple actions like walking, running to more complex interactions including shaking hands, answering the phone or drinking water. These actions or behaviors can get involve with one to multiple persons, objects, picturing a variable environment [2].

HAR is a field of computer vision and pattern recognition. It is an important area of computer vision research nowadays, it covers a broad scope of research topics such as human detection from video, human pose estimation, human tracking, and analysis and understanding of time series data [3]. Despite that, HAR is a challenging task due to the diversity of actions performed by humans in daily life. Several computer vision-based solutions that been proposed in literature were not necessarily successful due to the large video sequences that need to be processed in surveillance systems [4]. Even though, HAR has drawn much attention around the globe due to its promising results [5]. The primary challenge in HAR resides in how to design a powerful human action representation that is descriptive enough and computationally efficient [2].

HAR allow the enhancement of a wide range of human applications such as surveillance, telehealth, biometrics, video indexing, training or virtual coaching [6], Human Computer Interaction (HCI), ambient assisted living, human-robot interaction, entertainment, and content-based video search [5]. In HCI the activity recognition system observes the task undertaken by the user and then direct him\her towards its completion by providing feedback. In surveillance video systems, the activity recognition system can detect without manual intervention any unusual behavior and report it to the authorities for quick response. In the same way within the entertainment industry, these systems are capable to recognize and identify the actions and movements of different participants and multiple players during gameplay. Considering the complexity and time span, activities are categorized into four groups, i.e., gestures, actions, interactions, and group activities [7] [5]. The table below explain these types of activities. Figure2 and Table I show categorization of different level of activities [5].

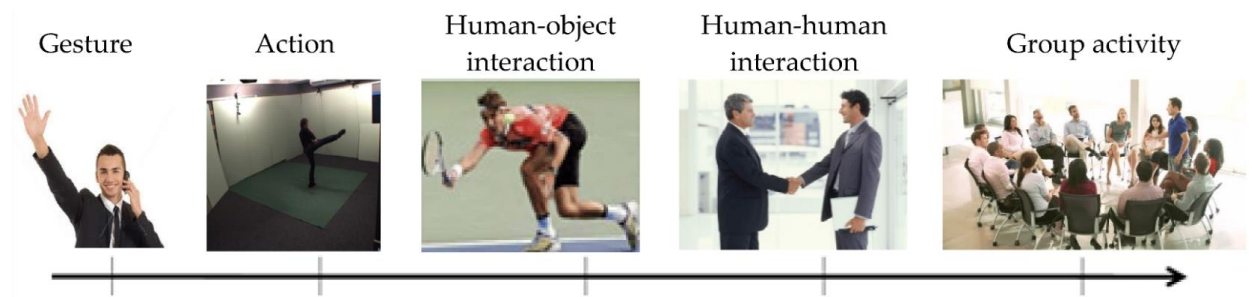


Figure 2:categorization of different level of activities

Table 1: categorization of different level of activities

	Definition	Examples
Gestures	the use of motions of the limbs or body as a means of expression [8]	raised his hand overhead in a gesture of triumph
Actions	Human motion in which the accomplishment of a thing is usually over a period of time, in stages, or with the possibility of repetition. It is the manner or method of performing. A function of the body or one of its parts. [8]	Basketball player dribbles, shoots, or passes the ball, considering both the function of the body and the specific parts involved in these actions.
Interactions	Human motion when two or more people or things communicate with or react to each other [9]	Shaking hands Playing with sand
Group activities	type of human motion that is composed of a sequence of actions [1]	group exercise class soccer game

Ultimately, HAR seeks to detect the person executing the action, or multiple persons executing activities on an unknown video sequence, identify how long the action lasts and classify its type. This is a challenging task that include identifying the person and the location and timing of the action performed by the person within the video stream, and in the end, action recognition and classification [1]. The most important application of action recognition is video surveillance [10]. Figure3 below shows some more applications.

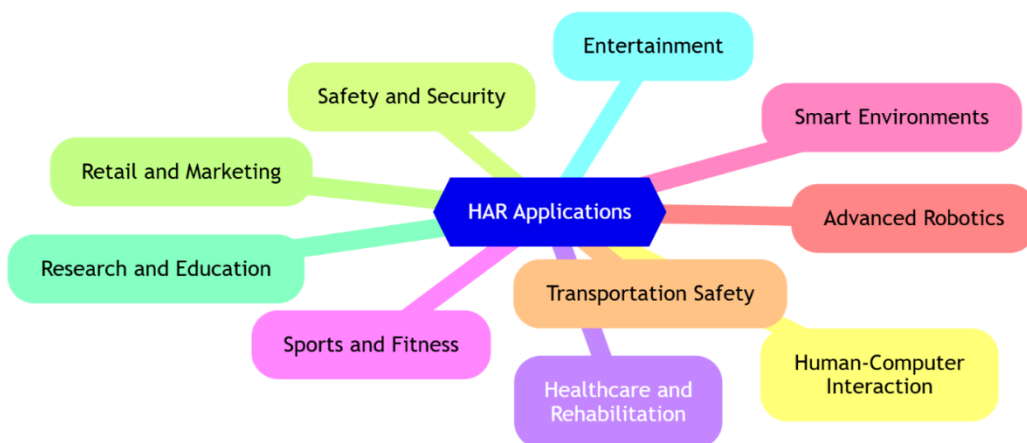


Figure 3: Application of HAR

2.1. Data Modalities

In HAR, various data modalities play a crucial role, each offering distinct advantages depending on the application. Commonly used modalities include RGB data, which provides rich appearance information but it is sensitive to background and lighting conditions, and skeleton data, which captures 3D structural information of human joints, offering robustness against background clutter and viewpoint changes. Depth data provides geometric shape information, useful in scenarios where color and texture are less important, while infrared data excels in low-light conditions but lacks color details. Non-visual modalities like audio can capture temporal sequences, acceleration data is valuable for fine-grained action recognition, and radar and Wi-Fi signals are used for privacy-preserving applications, especially in through-wall HAR. Combining these modalities through fusion-based or co-learning-based approaches can enhance HAR systems' accuracy and flexibility by taking advantage of the strengths of each modality. The figure below displays various data types used in Human Action Recognition, highlighting their roles in accurately identifying human actions [11], see figure4.







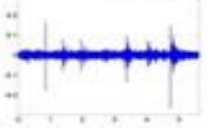
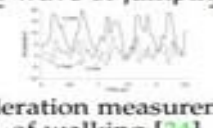
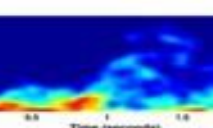

Modality		Example	Pros	Cons
Visual Modality	RGB	 Hand-waving [27]	<ul style="list-style-type: none"> Provide rich appearance information Easy to obtain and operate Wide range of applications 	<ul style="list-style-type: none"> Sensitive to viewpoint Sensitive to background Sensitive to illumination
	3D Skeleton	 Looking at watch [28]	<ul style="list-style-type: none"> Provide 3D structural information of subject pose Simple yet informative Insensitive to viewpoint Insensitive to background 	<ul style="list-style-type: none"> Lack of appearance information Lack of detailed shape information Noisy
	Depth	 Mopping floor [29]	<ul style="list-style-type: none"> Provide 3D structural information Provide geometric shape information 	<ul style="list-style-type: none"> Lack of color and texture information Limited workable distance
	Infrared Sequence	 Pushing [30]	<ul style="list-style-type: none"> Workable in dark environments 	<ul style="list-style-type: none"> Lack of color and texture information Susceptible to sunlight
	Point Cloud	 Bending over [31]	<ul style="list-style-type: none"> Provide 3D information Provide geometric shape information Insensitive to viewpoint 	<ul style="list-style-type: none"> Lack of color and texture information High computational complexity
	Event Stream	 Running [32]	<ul style="list-style-type: none"> Avoid much visual redundancy High dynamic range No motion blur 	<ul style="list-style-type: none"> Asynchronous output Spatio-temporally sparse Capturing device is relatively expensive
Non-visual Modality	Audio	 Audio wave of jumping [33]	<ul style="list-style-type: none"> Easy to locate actions in temporal sequence 	<ul style="list-style-type: none"> Lack of appearance information
	Acceleration	 Acceleration measurements of walking [34]	<ul style="list-style-type: none"> Can be used for fine-grained HAR Privacy protecting Low cost 	<ul style="list-style-type: none"> Lack of appearance information Capturing device needs to be carried by subject
	Radar	 Spectrogram of falling [35]	<ul style="list-style-type: none"> Can be used for through-wall HAR Insensitive to illumination Insensitive to weather Privacy protecting 	<ul style="list-style-type: none"> Lack of appearance information Capturing device is relatively expensive
	WiFi	 CSI waveform of falling [35]	<ul style="list-style-type: none"> Simple and convenient Privacy protecting Low cost 	<ul style="list-style-type: none"> Lack of appearance information Sensitive to environments Noisy

Figure 4: Different Data Modalities [11]

2.2. History of Human Action Recognition

The early HAR research studies in 1980s-1990s focused on recognizing human actions from images and videos via handcrafted features including motion trajectories, silhouettes, Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and optical flow to identify actions. Such approaches were somewhat elementary and often faced challenges with background clutter and occlusion [5]. Throughout the 1990s, sensor-based methods emerged. Technologies such as badge based tracking systems and acceleration sensors were employed to detect and recognize activities [12]. These approaches were especially effective in controlled environments such as office scenarios, where sensors could be strategically placed. Active badge systems were among the first to recognize activities for multiple users. The focus shifted to machine learning in the 2000s, traditional machine learning algorithms such as SVMs (Support Vector Machines) together with handcrafted features. Temporal models including HMMs (Hidden Markov Models) and HCRFs (Hidden Conditional Random Fields) [13], were employed to capture the sequential nature of actions. The impact of deep learning emerged in 2010s. Approaches like CNNs (Convolutional Neural Networks) and RNNs (Recurrent Neural Networks) revolutionized HAR. These models capture multi-level features from raw data and had markedly increased recognition accuracy [14]. In the 2020s, HAR continues to evolve with the current studies which directed towards multimodal fusion, spatiotemporal modeling and self-supervised learning. Even with the progress, issues like addressing occlusions, changing in camera angle -viewpoint- and lighting variations continue to be major obstacles that researchers seek to resolve actively [15]. See figure 5 for a visual summerization.

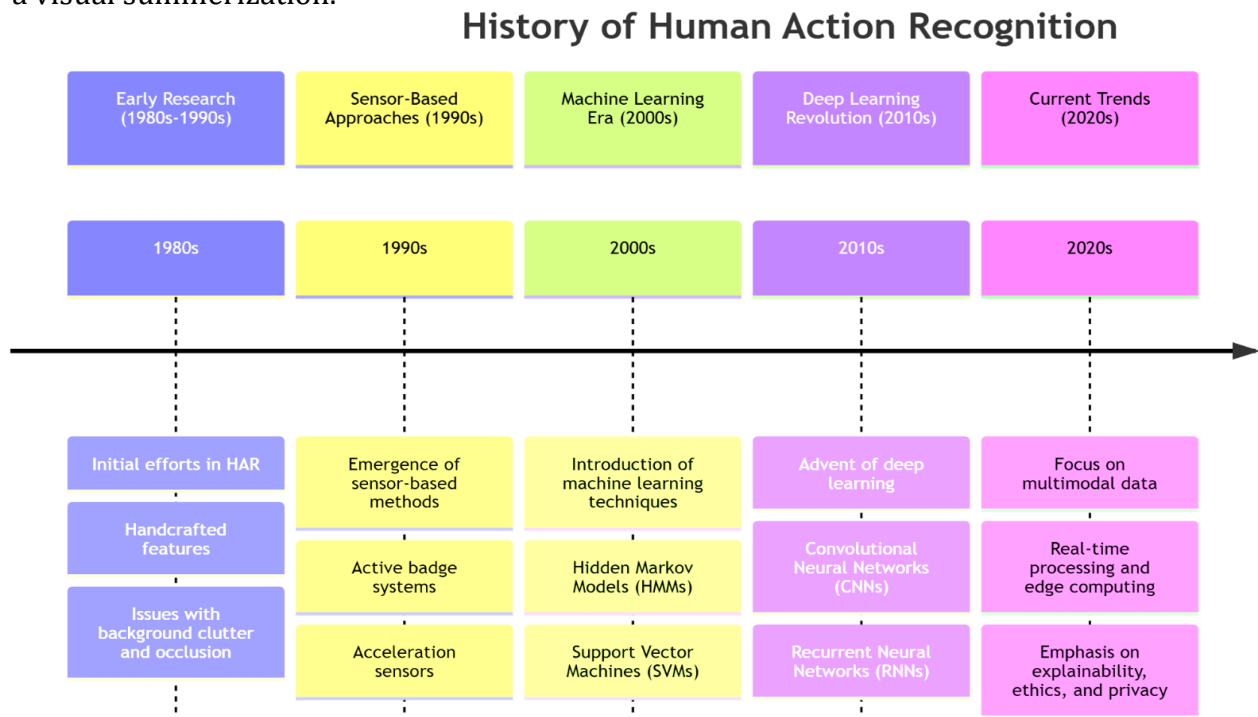


Figure 5: History of HAR

2.3. Data collection for HAR tasks

The following table contains set of datasets that have been used in HAR tasks, and figure6 shows some examples of human action categories across various widely used datasets.

Table II Data sets used in HAR tasks

Dataset	Data Collection Method	Annotation Process	Challenges Faced
UCF101 [16]	Downloaded from YouTube, Large number of action classes	Manually labeled with a single action class for each video clip	Unconstrained videos from YouTube present several challenges, including varying quality, camera motion such as panning and zooming, and diverse lighting conditions. Additionally, these videos often feature partial occlusion of objects or people and low-quality frames. The complexities within the actions themselves add to the difficulty, with informative motions often occupying only a small portion of the clip.
Kinetics [17]	Downloaded from YouTube, focusing on a wide range of human actions.	Annotated using crowdsourcing platforms like Amazon Mechanical Turk.	Handling variations in video quality and ensuring accurate annotations.
HMDB [18]	Videos from various movies, public databases, and user-generated content.	Manually labeled, with a focus on motion and action classes.	Managing heterogeneous video sources and achieving consistent labeling.
ActivityNet [19]	Videos collected from YouTube, covering a diverse set of activities.	Annotated using a hierarchical labeling system with multiple annotators for verification.	Deals with long video durations.
AVA [20]	Movie clips (15 minutes each) collected and annotated for spatial-temporal actions.	Annotated manually, focusing on atomic visual actions with precise spatial and temporal localization.	Handle complex scenes.

Something-Something [21]	Videos collected by asking participants to perform specific actions with objects.	Annotated manually with fine-grained action labels, ensuring temporal consistency.	Ensuring uniformity in action performance and capturing subtle variations.
--------------------------	---	--	--

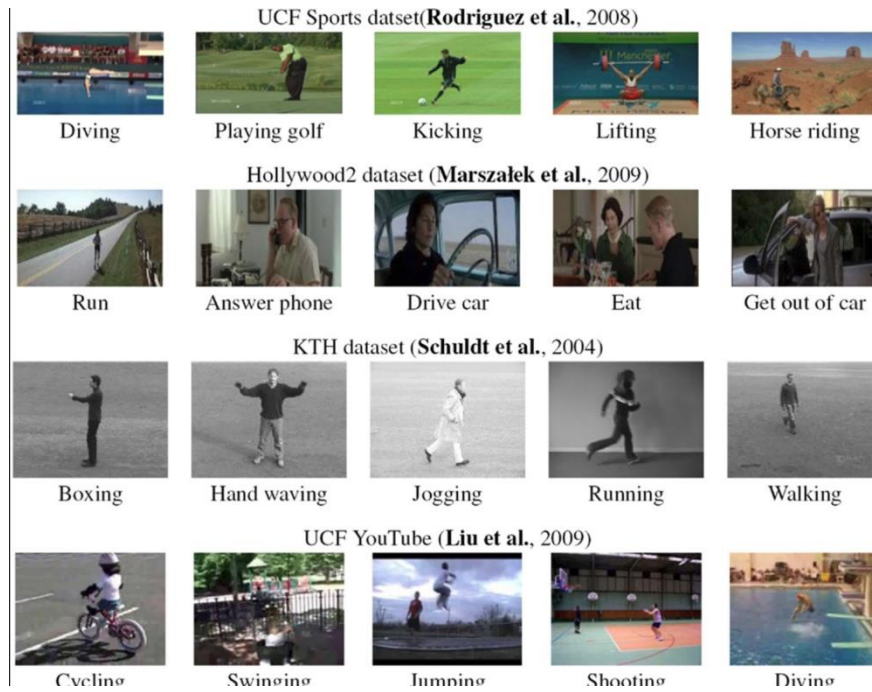


Figure 6: shows some examples of human action categories across various widely used datasets, featuring UCF Sports, Hollywood2, KTH, and UCF YouTube, highlighting the diversity in actions and environments used for benchmarking human action recognition models

2.4. Core Components of HAR

In Human Action Recognition (HAR), the workflow begins with data acquisition, a crucial step involving the collection of data from various sources such as video cameras, depth sensors, and wearable devices. This data may include a range of information like visuals, sounds, and physiological measurements [13]. Once collected, the data is labeled with action tags and subjected to preprocessing techniques such as normalization and data

augmentation [22]. These steps ensure that the dataset is consistent and diverse, which is essential for building robust recognition models.

Following data acquisition, feature extraction transforms this raw data into meaningful attributes that represent human actions. This process involves identifying important spatial and temporal details such as shapes, textures, and body movements [23]. Various techniques, including pose estimation and deep learning models like Convolutional Neural Networks (CNNs), are employed to gather these features [22]. Motion features may also be directly extracted from video frames using filters that highlight recent movements, avoiding the need for complex tracking methods [23]. Techniques like Histogram of Oriented Gradients (HOG) and Speeded Up Robust Features (SURF) are commonly used to capture critical aspects of human movement [13]. To manage the complexity of the data, methods such as Principal Component Analysis (PCA) are applied to reduce dimensionality while retaining essential information [23].

The final step involves classification and recognition, where the extracted features are used to identify and categorize actions. Algorithms such as Support Vector Machines (SVM), Hidden Markov Models (HMM), and advanced deep learning models like CNNs and Recurrent Neural Networks (RNNs) analyze these features to predict the action class in new data [13]. This process not only classifies the action but also locates it within the data, which is important for applications like video surveillance and human-computer interaction [22]. The performance of HAR systems depends on the quality of the extracted features, the choice of algorithms, and the system's ability to handle various conditions, including changes in scale, viewpoint, and occlusions. The success of these systems is evaluated based on metrics such as accuracy, precision, and recall, reflecting how well they generalize to different scenarios [13].

2.5. HAR Current Trends

The research field is going towards some recent trends that are making a revolution in the human action recognition field. Hybrid models is one of these trends, Which is a combination of different techniques in order to improve the performance [24], one of the novel combination for a hybrid model was in [25], which combined CNN and a transformer called Vision Transformer ViT and it showed a great performance with an accuracy of 97.89%, which is high compared with using only CNN or ViT. Visual transformers are gaining attention for their effectiveness in capturing long-range dependencies in video data [26]. there is also new trend of using reinforcement-based learning in HAR, which is possible to have a system that are capable of learning themselves from punishment and reward schemes, such as in [27].

2.6. Challenges of HAR

Human action recognition is not a simple task, it has a lot of obstacles that make it complex. The main obstacle that HAR had in its complexity form is the complexity of the human actions

itself, and the wide ways of representing it. Additionally, the intra class variability is a problem, which is the variety of the ways of doing the same action.

Complex backgrounds can also cause a problem in the process of extracting the features. The data itself can cause a problem when there are different viewpoints (see figure7) or different distances from the camera or even when it's noisy. And when it comes to videos, Actions unfold over time, involving complex temporal dependencies, therefore recognizing actions requires capturing both short-term and long-term dynamics. The choice of the classification algorithm can be challenging depending on the type of data and other factors [11] [6] [28] .

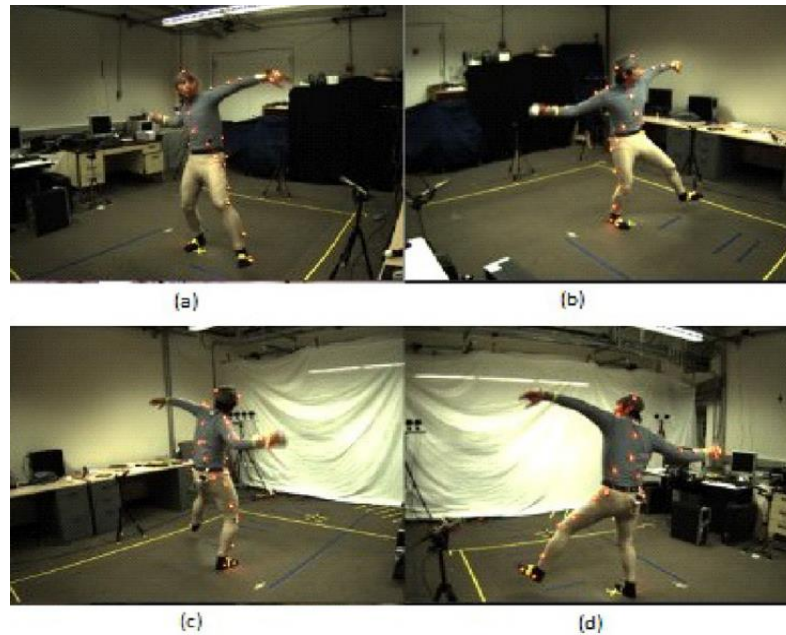


Figure 7: A person performing the same action “throwing” from multiple views. Images from Berkley MHAD dataset [29]

2.7. Future Work

HAR is a rapidly advancing field, and the future is holding a massive evolution in our interaction with technology. The advancement in HAR won't only revolutionize applications in healthcare, sports, and security but also in our daily lives. The future direction in HAR is going towards multimodal approaches that can combine information from multiple resources (such as RGB, Depth, and Skeleton data), also it is showing interest in refining the deep learning methods used in HAR, which has already shown significant successes for this domain. Developing context-aware models that are capable of understanding sequences of activities could be of great help in further improvement of action recognition accuracy. The privacy of the user is being the main concern to achieve in any vision-based systems [15] [30].

3. HAR in sports

Human Action Recognition (HAR) has a significant impact in a wide range of sports encompassing soccer [31] [32] [33], volleyball [34] [35], basketball [36] [37] [38], weightlifting [39], swimming [40] [41] [42], tennis [43] [44] [45], boxing [46] [47] [48], wrestling [49] [50], hockey [51] [52], and dancing [53] [54]. HAR systems are used to detect and recognize actions and activities of players and teams through different stages such as training, matches, warm-ups, or competitions [55]. The main goal of HAR is to pinpoint the athlete carrying out the action, figure out the duration of the action and then classify the action type. In sports context, HAR is key to track athlete performance. Allows for spotting and following motions, specific action recognition, diverse techniques analysis, evaluation of skill levels, and automatic data analysis support. The variety of actions in sports can be broad and complicated, including movements carried out by players to accomplish specific tasks, commonly involving collaboration with objects or other team members. Taking into account the divers forms of actions in sports, it is important to classify these actions by utilizing an organized framework considering their complexity, performance standards, and type of interaction. This organized framework helps in getting deeper understanding and optimizes the effectiveness of action recognition systems across multiple sports. Research have revealed how HAR systems can be customized to fulfill the specific needs of different sports. Whether it be in team sports, individual sports, combat sports, athletics, or performing arts as figure8 shows below.

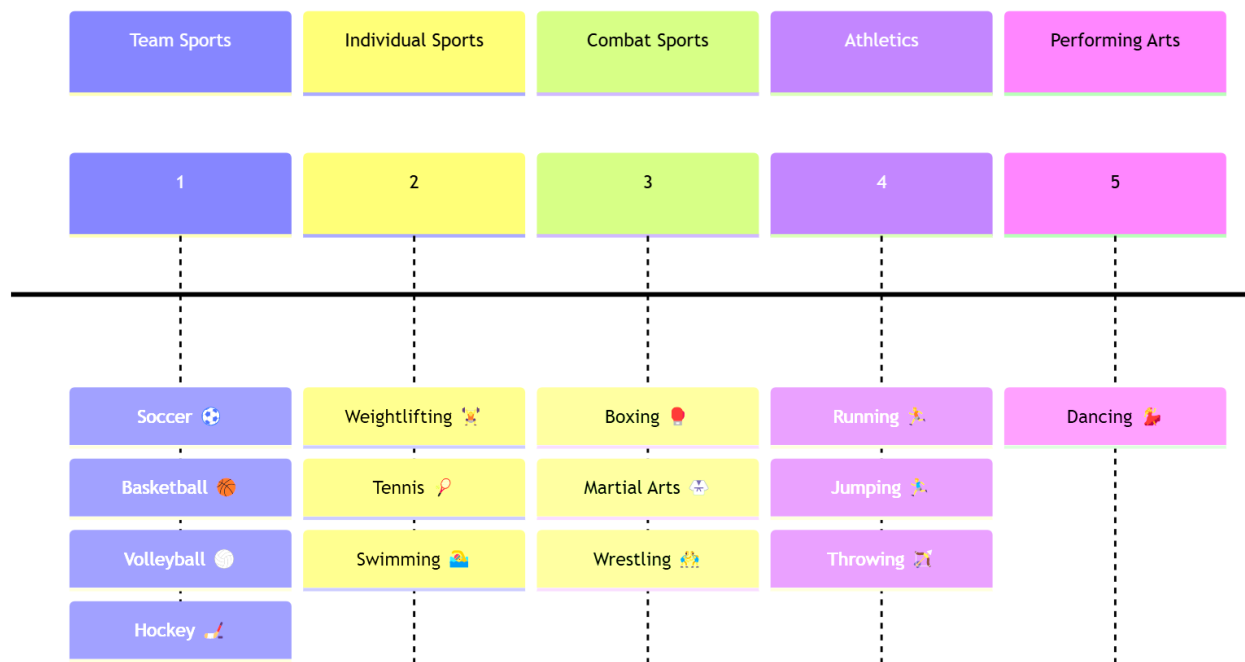


Figure 8: different types of sports

3.1. Applications of HAR in sports

HAR in sports is a game-changer, encouraging innovation across different athletic domains. From enhancing player performance to optimizing training routines, variety of HAR's applications are altering the sports industry, this section will mention a set of these applications.

Injury Prevention in HAR systems reliably track and analyze athlete movements throughout training sessions and live games. These systems provide key insights towards athlete physical movements which allow coaches to point strengths, limitations and areas to improve training programs and deliver personalized recommendations to strengthen techniques and general performance. Injuries sustained during sports of professional athletes limit the enhancement of sports performance. With this understanding, an effective and accurate systems for assessing sports injuries is necessary to detect possible injuries in athletes and to implement injury prevention programs [56]. By analyzing movement patterns, such systems can spot potential injuries before they become severe. This forward-thinking approach contributes to keep support on athlete health and also to secures continuous performance enhancement without injury-related disruptions [56] [57].

Performance Analysis and Improvement involve the evaluation of an athlete movements and skills to find improvement areas leading to the improvement of the athletic performance. Applying technologies like video image processing and optimization algorithms such as PSO (particle swarm optimization algorithm), coaches and athletes can precisely detect variations from standard techniques [58], analyze the succession of current training programs and then make analytics-based adjustments. This help in correcting technical flaws.

Player Scouting and Recruitment recognize and appraise potential talent for sports team. By utilizing techniques like video analysis and data-driven (analytics-based) algorithms. Talent scouts can evaluate a player's capabilities and talents, reliability in performance, and how well they fit within the team. These techniques enable more unbiased and comprehensive assessments, assisting teams in making data-driven decisions when choosing players who fulfill the team requirements and show potential to perform well at higher level competitions. This technique increases the precision and effectiveness of the scouting (talent identification) and requitement (selection) process [59] [60]. A talent identification in youth ice hockey that explore intangible player characteristics [61].

capturing special actions frames provided an assistance to all news companies and the crowd who is interested in knowing only the key moment of the game or a quick summary of the less or more 90 minutes match , therefore such an application were introduced in [62], which a system made up of deep learning-based action recognition, object identification, and facial recognition components. The system uses the previously mentioned method to

highlight the key frames of the match, then a short stunning film is produced by merging the highlighted pictures and frames of the special moments. The main keyframes were featured as shot segmentation, red and yellow card recognition, corner kick detection, penalty kick detection, shoot and celebration detection, score detection, and face identification (see figure9).



Figure 9: capturing free throw as a keyframe [62]

3.2. Sports

Sports provide a dynamic and challenging domain for research on Human Action Recognition (HAR). Each type of sport brings in a unique set of movements and interactions, which push the boundaries of HAR technology. The following sections discuss how HAR is applied in each sport, highlighting major challenges and innovations in these areas.

3.2.1 Basketball

Basketball is a high-energy sport known for its complex plays. Human Action Recognition (HAR) in basketball involves identifying and analyzing the complex movements and interactions of players on the court.

A. Approaches

In the evolving field of human action recognition (HAR) in basketball, researchers are using advanced algorithms to better analyze players' movements (see figure11). Deep learning techniques, such as convolutional neural networks (CNNs) and bidirectional Long Short-Term Memory (Bi-LSTM) networks, are effective in identifying actions from data collected through sensors and video [63]. Other methods like the Lucas-Kanade algorithm for optical

flow, Gaussian Mixture Models (GMM) for motion descriptors, and Linear Discriminant Analysis (LDA) for feature representation are also useful [64]. Combining data from multiple sensors and improving data fusion methods can further enhance results [65].

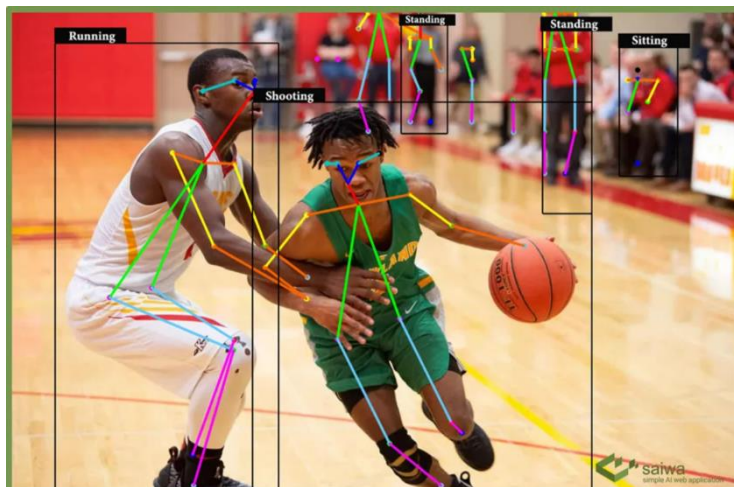


Figure 10: This figure illustrates various player movements and positions during a basketball game. The player labeled 'Shooting' is in the act of taking a shot, while another player labeled 'Running' is moving towards them. Additional labels indicate crowd actions such as sitting and standing

Techniques like Part-aware Long Short-Term Memory (PLSTM) networks and real-time processing with Support Vector Machines (SVM) are promising for faster and more efficient recognition [65]. For detailed action recognition, methods such as Two-Stream Networks, Temporal Segment Networks (TSN), and Inflated 3D ConvNets (I3D) work well, especially when combined with frameworks like Navigator-Teacher-Scrutinizer and the fusion of RGB and optical flow data [66]. In recognizing basketball stances, the PCA+CNN algorithm, which uses Principal Component Analysis for better feature extraction, has shown high accuracy [67]. Future research should aim to combine different types of data, use ensemble methods, add synthetic data to training sets, improve model structures, and use advanced techniques like transfer learning to boost accuracy and reliability. This will benefit real-time applications in training and performance analysis for athletes and coaches.

B. Datasets

SpaceJam [68] is a dataset of 32560 example contains 9 action categories capturing single player actions in basketball including step, race, block, dribble, ball in hand, shooting, position, walk, defensive position and one category with no action label. *SpaceJam* can be

used for player tracking, pose estimation, sports analytics, AR/VR applications, robotics, health and injury prevention.

Hang-Time [69] collected from 24 players using wrist-worn inertial sensors from both basketball training sessions and full games, players where from USA and Germany. These sensors capture data such as acceleration, gyroscope readings, and possibly magnetometer data, which are used to monitor and recognize different physical activities performed by basketball players during training sessions and full games as mentioned.

KTH [64] composed of 599 video clips from 25 individuals from four different scenarios. this dataset consists of six human actions including boxing, handclapping, handwaving, jogging, running, and walking.

UCF [64] dataset consists of 150 videos of ten different activities of human body motion. The activities include running, walking, jumping, and other simple actions. The UCF datasets are known for their challenging conditions, including lighting variations, camera angles, and occlusions. (See figure12).

BARD [64] a basketball action recognition dataset collected from 3833 YouTube video clips which lasts about 10 sec on average. these clips include actions like shooting, passing, jumping, defending, and running. The challenge with BARD was the similarity between actions like shooting and jumping which make the accurate recognition difficult.

The dataset mentioned in the paper titled "Human Motion Recognition by Three-view Kinect Sensors in Virtual Basketball Training" [65] include actions like standing, dribbling ball, shooting ball, and passing ball, actions to represent the motion used in VR basketball application. Twenty participants of different body size, gender, and level of experience in playing basketball were requested to perform each of these actions 30 times, resulting in around 130,000 labeled frames in the dataset. This was captured using three Kinect sensors positioned around the subject in a 360-degree scenario to ensure maximum views of all user's actions. 3D skeleton data were collected at a speed of 30 frames per second with the help of the RGB camera and infrared emitter of the Kinect sensor, giving detailed information about the positions and movements of all 25 joints in a human body.

A basketball dataset is proposed in the paper "Fine-Grained Action Recognition on a Novel Basketball Dataset" [66] contain broadcast videos of basketball games captured from different perspectives with various camera movements with a total duration of 8 hours for all. The dataset consist of 3399 annotated instances of three broad actions that including dribbling, passing, and shooting which are further categorized into 26 fine-grained actions. It have some challenges due to unrelated objects as well as complicated backgrounds, the temporal and spatial scales of the actions are small, the interclass difference is minor, and the occlusions are frequent. Skeleton-based action recognition can be used if RGB frames are utilized to extract skeleton data using pose estimation.



Figure 4: sample of basketball actions frame from UCF dataset

C. Challenges

Due to the complex actions in basketball match and the fast movement [67], accurate HAR in basketball were difficult. This section explores the key obstacles faced by the HAR researchers in basketball.

One of the notable challenges is the high similarity between various actions in basketball that may lead to inaccurate performance in detecting specific movements [64] [65] [63]. And the common complex backgrounds can be a problem for some approaches [66] [64].

Lighting conditions also can play a critical role in affecting the performance of the classifiers [70] [64]. Dynamically changing body coordinates impact the accuracy of research that used the skeleton data modality, such as the inaccurate detection of the player's back and front [65] [71]

Besides, the correlation between the features and the feature dimensionality can strongly influence the classification efficiency as highlighted in in [67]. Lastly, there is a notable gap in the HAR in basketball using wearable sensors. This gap affects the prior knowledge, such as the best choice of type of wearable sensors and how to optimize their usage including managing power consumption [67] [63].

3.2.2. Football

Football, the world's most popular sport, engages millions with its dynamic and fast-moving gameplay. Human Action Recognition (HAR) in football seeks to analyze and understand the diverse movements and strategies employed by players on the field.

A. Approaches

Using a traditional approach where features manually extracted by taking the absolute sum of accelerometer or gyroscope signals during the kick phase then fed the features into SVM classifier to recognize shots, passes and null class [72]. The SVM to succeed, the quality of the manually engineered features need to be high. An improved dense trajectory IDT is a manual feature extraction method used to isolate players movements from any camera-induced movements, this method used to extract features like trajectory shape, histogram of oriented gradients HOG, histogram of optical flow HOF, and motion boundary histograms MBH, principal component analysis PCA then used to reduce the dimensionality of the feature space [73] [74] and reduce noise [74]. Optical flow used to extract temporal features and to capture dynamic aspects of actions [75].

To automatically extract features from images or video frames a CNN used for both lower level and high-level features to be extracted [76] [73] [75].

The extracted features combined with meta information (player and ball location) to form a person centered features [76], LSTM networks used for temporal modeling and classification [76] [72] a combined convolutional and LSTM additionally used [41] allowing the model to extract specific movements features and temporal features to show how these movements progress over time. The complexity of using CNNs and LSTM along with multi camera setup [76] or with convLSTMs [72] is very high and need to be optimized within modern hardware GPUs or edge devices to be run on real time, that indicate that there is a real time implementation challenges [76] [72] .

The CNN used [75] is an upgrade version with the inclusion of batch normalization layers into a model name BN-Inception model. This model makes the build of deeper networks without facing issues like vanishing gradients more efficient. The model captures the static aspects of the scene and combined with dynamic feature extracted from optical flow to be integrated to fusion methods including average fusion, max fusion and fully connected fusion. The paper achieved accuracies are 95.2% on the UCF101 dataset and 73.4% on the HMDB51 dataset using fully connected multimodal fusion. The gap between accuracies on two different datasets show that the approach may not fully represent the diversity of action recognition scenarios.

CNN in [73] used to extract spatial features from video frames while processing into two dimensions which are time and RGB channels. The temporal dimension captures the dynamics motion over time and the RGB channels identify color-based features that help to distinguish different actions [73]. A hybrid approach used to extract features by combining the IDT and the CNN through a fusion process [73] , these extracted features fed into KELM classifier which is a type of FNN with higher learning speed and generalization performance the hybrid approach introduce complexity that effect the scalability, real time applicability and the ease of integration [73]. A multi-resolution 3-dimensional convolutional neural

network is constructed by combining the convolutional neural network and a 3-dimensional neural network.

B. Datasets

FIFA were the first resource to think of when talking about football , the datasets in this case will be consistent of footages from matches labeled with set of actions or even a video records from the *FIFA* games as in [77] [78] [62] . Another dataset used in this field is *SoccerDB1* such as in [79][Novel Architecture for Reducing Sports Injuries in Football], which it is consistent of 448 soccer video clips labeled in 4 actions classes (Dribble, Kick, Run, and Walk), with 70 clips for each class. *SoccerNet-v2* is a remarkable dataset that consistent of 550 matches labeled with 17 general action, [80] [81] used it in their research , in fact *SoccerNet-v2* is an enriched version of *SoccerNet* which was only consistent of 3 actions labels (goal, card, and substitution) [82] [83] .*UCF101* and *HMDB51* are datasets that share some common features. *HMDB51* contains 11 action categories, such as passing, heading, and chasing, while *UCF101* includes 50 action categories , at least one of these two datasets where used in [84] [85] [86] [87] [88].

C. Challenges

Recognizing human actions in football videos is challenging for several reasons. Complex backgrounds and changing lighting conditions can make it hard to see players' actions clearly [74]. Different camera angles can lead to misinterpreted movements [74]. Building deep learning models for sports action recognition requires lots of high-quality labeled data, which is hard to get due to different match conditions and player behaviors [62]. Real-time recognition during live broadcasts is difficult because it needs fast processing and good resource management to avoid delays [62]. It's also tough to distinguish between similar actions like passing and shooting and to handle the movement of multiple players and the ball [76]. Large models with many parameters use a lot of memory and are slow to train, making them hard to manage [75]. Recognizing group actions is even harder due to the focus on individual actions and the complications from camera movement and video quality [89]. The lack of annotated datasets for complex actions like dribbling makes training and validation difficult [90]. Real-time analysis also needs efficient algorithms and high computational power to keep up with the fast pace of soccer matches [90].

3.2.3. Tennis

Tennis is a thrilling sport characterized by rapid rallies, powerful and strategic gameplay. Human Action Recognition (HAR) in tennis focuses on capturing and analyzing the precise

and varied movements of players, offering insights into their techniques and strategies on the court.

A. Approaches

In HAR the workflow is similar even in different tasks, starting with feature extraction then actions classification. In the task of tennis, the same workflow is followed.

In vision-based data, due to its remarkable success in the image classification field [91], CNN is the most popular and used method for learning complicated features [92], as in [93], an inception architecture of CNN was used, which is a network of 22 traditional convolutional layers stacked in lower layers and inception modules stacked at the higher layers. Also [94] used CNN in their work, and in [95] CNN were used to extract 2 action labels from the images that were fed to the network.

While in sensor-based data such as in [96], raw data were simply used after performing normalizing and rescaling.

For the next step in the workflow, which is the classification part, a variation of classifiers can be noticed. same papers go towards common classifiers, as how [96] used the support vector machines (SVM) and K-nearest neighbor (KNN) classifiers, in SVM the RBF-kernel were used, and the similarity measure in KNN. [97] also employed SVM in their work because of its advantages of providing better prediction of unseen test data, providing a unique optimal solution for a training problem and containing fewer parameters than other parameters. additionally, 12 non-linear SVMs were used in [98]

Long short-term memory (LSTM) is also a common classifier used in action recognition due to their power in modeling the dynamics and dependencies in sequential data [99]

[deep learning for domain-specific action recognition in tennis] used 3 stacked layers of LSTMs to give the best results. As LSTM works efficiently with sequence or time-series data [100], there are other methods that can be used with this kind of data as proposed in [94], dynamic time wrapping was used which uses the dynamic programming algorithm to find the optimal matching of two sequences.

The deep learning classifiers are also commonly used in action recognition tasks. The method in [101] were based on the deep learning classifiers called graph convolutional network GCN, they integrated the attention mechanism and ST-GCN, which extends GCN to the space-time domain, this method has achieved promising results in the action recognition domain [102].

B. Datasets

The THETIS dataset is crucial for fine-grained tennis action recognition, featuring 1980 RGB videos of 12 different actions performed by 55 players. This dataset supports detailed classifications for both training and performance analysis, reflecting the complexities of real-world scenarios [93]. In a related study, data was collected from eight amateur tennis players, including both men and women, capturing various tennis strokes (forehand,

backhand, volley, lob) along with everyday activities (walking, running, jumping). This dataset includes 577,097 labeled samples from 12 sensor axes (3 accelerometers and 3 gyroscopes), leading to 144,201 spectrograms for analysis. The classification system achieved impressive accuracy, with 99.25% for known players and 96.51% for new players, demonstrating its effectiveness in recognizing tennis strokes despite player differences [103]. Additionally, another dataset comprised 769 tennis strokes—212 backhands, 197 forehands, and 360 volleys—using player silhouettes and rackets. This dataset was divided into training (60%), validation (20%), and test sets (20%), with results showing that incorporating both racket and silhouette data improved classification accuracy, achieving an RMSE of 5.31% compared to 9.64% for silhouettes alone [45]. This comprehensive approach highlights the dataset's applicability and accuracy in tennis action recognition.

C. Challenges

There is a general lack of publicly available, standardized benchmark datasets for fine-grained action recognition in tennis, which constrains the possibility of comparing and validating their results against other methods. [93]Dynamic environments like sports arise camera challenges to recognize actions constantly. The methodology [96]is based on visual data, and all types of visual data are extremely sensitive to the camera point from which they were captured. If the test camera viewpoint is far from the train camera viewpoint, recognition accuracy drastically falls which affects robustness in the system across a variety of scenarios .Taking this into account sensors placed on different body parts effect the accuracy of the collected data due to different environment that introduce variability like lighting, occlusions, or outdoor versus indoor settings [96]To address this challenge that make the system less sensitive to camera angles one may use multiple cameras ,creating features that work well from any angle may be considered .the placement of the sensor on the best part of the body to get the best data come by experiments .improving the system to handle different light conditions and occlusions [104] .The accelerometers used are prototypes which need to be smaller and less obtrusive to satisfy players comfort during performance .There is limited action classes of tennis strokes (serve, forehand , backhand) which do not cover all actions in areal match [96].

The task proposed by [93]is to recognize specific tennis actions, like forehand and backhand, that have high intra class variability, for example, different styles of forehand, and pretty low inter class variability, for example, between a forehand and backhand can often get misclassified by the model as similar actions. THESIS dataset used [93]contain low resolution and single camera videos with dynamic background and frequent occlusions and it is a small dataset that may be risky to overfitting. Due to the temporal dependencies between frames in the video data. it is challenging to model and extract meaningful features that capture the dynamics of the actions. Furthermore, the tennis ball does not appear on the dataset videos which result a difficulty on distinguishing between actions like different type of serves or between serve and a smash. More often, the model's errors were interpretable, such as confusing similar types of tennis strokes. This indicates that while it learns

semantically meaningful information, it gets confused about the fine differences between some of the actions. Performance of the model varied when classifying actions from amateur versus professional players, showing the fact that the ability of the players is going to impact the model in recognizing actions correctly. This suggests that the model may need to learn different features. [93]

The collection of data on Inertial Measuring Units (IMU) from wearable devices and correlating this data with labelled psychological states by expert coaches was challenging because the labels depended on the subjective observation of the coaches making the labels potentially inconsistent and hard to standardize; the small sample size of participants—who were just four players and two coaches—limited the generalizability of the findings. Machine learning models trained using one player's data did not generalize to other players very well. This is another proof of the case where psychological states, especially the "zone," are very personal, hence demanding personalized models per player. This makes it hard to build a one-size-fits-all solution and adds extra data collection and model training in the case of a new player. Finally, because of the use of off-the-shelf wearable sensors, this constrained the type of data that could be collected. Only inertial measurement unit data were used, which excluded potentially very useful physiological data. There were some practical difficulties associated with experimenting on elite athletes in the real world such as how to collect data accurately and consistently without interfering with the athletes during performance.

4. Conclusion and Discussion

Human Action Recognition (HAR) in sports in this review highlights a significant advancement within both methodologies and applications presented. To recognize and analyze complex human actions in sports including basketball, football, and tennis. Many approaches have been employed. Starting with traditional methods, such as Support Vector Machines (SVM) and manual feature extraction, these methods have been useful in less complex scenarios but often fall short when dealing with the complex and dynamic characteristics of sports actions. In comparison with deep learning techniques, especially CNNs and LSTM networks that have delivered outstanding results in capturing both spatial and temporal features, producing more accurate recognition results.

While HAR is helpful for improving training, game strategies, and player safety, there are still some challenges in making sure methods are working well in all cases. Because of the variety between actions and the dynamic environment in the field of sports, this brings out the complex nature of these actions and introduce considerable challenges, such as background clutter, lighting variations and occlusions that add difficulties in making precise action detection and classification. High quality and well labeled datasets availability are another key challenge Especially in sports that involve broad range of actions. Above all the need of

efficient deep learning algorithms and advanced hardware to attain feasible performance in real time action recognition to it to become applicable in real world scenarios.

From our study, it's clear that there is no single method that works best for everything in sports. Choosing the right HAR approach depends on carefully considering the specific needs of the situation. Even though HAR has come a long way, some problems still need to be solved as the field continues to grow. By addressing previous challenges, integration of Transformer models into future research should be taken under consideration. This is because they have previously recorded superior performance in modeling long-range dependencies and contextual information in a wide range of domains, making them more effective at handling sequential data compared to traditional CNNs and LSTMs; hence, they remain a robust and scalable solution to complex tasks in HAR. This could make the transformers very useful additions to the existing methodologies if integrated into HAR, since it would increase recognition accuracy and the possibility of processing more complex and overlapping actions.

Future research needs to target the challenges addressed by exploring multimodal data fusion, strengthening model stability and robustness across varying scenarios, employing more detailed data sets collections. The integration of wearable sensors and the optimization of algorithms for real-time processing are also promising trends that could increase the HAR applicability in sports. Moreover, data privacy ethical concerns must not be ignored as this field is attracts more attention.

References

- [1] Host, K., & Ivašić-Kos, M. (2022). An overview of Human Action Recognition in sports based on Computer Vision. *Heliyon*, 8(6)..
- [2] Zhen, Zhen, X., & Shao, L. (1999). Introduction to human action recognition. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1-11..
- [3] Zhang, H. B., Zhang, Y. X., Zhong, B., Lei, Q., Yang, L., Du, J. X., & Chen, D. S. (2019). A comprehensive survey of vision-based human action recognition methods. *Sensors*, 19(5), 1005..
- [4] Khan, S., Khan, M. A., Alhaisoni, M., Tariq, U., Yong, H. S., Armghan, A., & Alenezi, F. (2021). Human action recognition: a paradigm of best deep learning features selection and serial based extended fusion. *Sensors*, 21(23), 7941..
- [5] Applied Sciences | Free Full-Text | A Comprehensive Review on Handcrafted and Learning-Based Action Representation Approaches for Human Activity Recognition (mdpi.com).
- [6] Ramanathan, M., Yau, W. Y., & Teoh, E. K. (2014). Human action recognition with video data: research and evaluation challenges. *IEEE Transactions on Human-Machine Systems*, 44(5), 650-663..
- [7] Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *Acm Computing Surveys (Csur)*, 43(3), 1-43..
- [8] Merriam-Webster. (n.d.). Merriam-Webster: America's most trusted dictionary. Merriam-Webster. <https://www.merriam-webster.com/>.
- [9] SharePoint. (n.d.). Intro to human action recognition [PowerPoint presentation]. SlideShare. <https://www.slideshare.net/human-action-recognition/intro-to-human-action-recognition-847021795>.
- [10] 5. Liu D., Xu H., Wang J., Lu Y., Kong J., Qi M. Adaptive Attention Memory Graph Convolutional Networks for Skeleton-Based Action Recognition. *Sensors*. 2021;21:6761. doi: 10.3390/s21206761. [PMC free article] [PubMed] [CrossRef] [Google Scholar] [Ref list.
- [11] "Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*, 45(3), 3200-3225."
- [12] Want R., Hopper A., Falcao V., Gibbons J.: The Active Badge Location System, *ACM Transactions on Information, Systems*, Vol. 40, No. 1, pp. 91–102, January 1992.
- [13] Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2, 28..

- [14] Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*, 45(3), 3200-3225..
- [15] Progress of Human Action Recognition Research in the Last Ten Years: A Comprehensive Survey | *Archives of Computational Methods in Engineering* (springer.com).
- [16] "Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*."
- [17] "Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., ... & Zisserman, A. (2017). The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*."
- [18] "Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011, November). HMDB: a large video database for human motion recognition. In *2011 International conference on computer vision* (pp. 2556-2563). IEEE."
- [19] "Caba Heilbron, F., Escorcia, V., Ghanem, B., & Carlos Niebles, J. (2015). Activitynet: A large-scale video benchmark for human activity understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 961-970)."
- [20] "Murray, N., Marchesotti, L., & Perronnin, F. (2012, June). AVA: A large-scale database for aesthetic visual analysis. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 2408-2415). IEEE."
- [21] "Goyal, R., Ebrahimi Kahou, S., Michalski, V., Materzynska, J., Westphal, S., Kim, H., ... & Memisevic, R. (2017). The "something something" video database for learning and evaluating visual common sense. In *Proceedings of the IEEE international conference*."
- [22] Yao, B., Jiang, X., Khosla, A., Lin, A. L., Guibas, L., & Fei-Fei, L. (2011, November). Human action recognition by learning bases of action attributes and parts. In *2011 International conference on computer vision* (pp. 1331-1338). IEEE..
- [23] Masoud, O., & Papanikolopoulos, N. (2003). A method for human action recognition. *Image and Vision Computing*, 21(8), 729-743..
- [24] Abbaspour, S., Fotouhi, F., Sedaghatbaf, A., Fotouhi, H., Vahabi, M., & Linden, M. (2020). A comparative analysis of hybrid deep learning models for human activity recognition. *Sensors*, 20(19), 5707..
- [25] Alomar, K., Aysel, H. I., & Cai, X. (2024). RNNs, CNNs and Transformers in Human Action Recognition: A Survey and A Hybrid Model. *arXiv preprint arXiv:2407.06162*..
- [26] Kulbacki, M., Segen, J., Chaczko, Z., Rozenblit, J. W., Kulbacki, M., Klempous, R., & Wojciechowski, K. (2023). Intelligent video analytics for human action recognition: the state of knowledge. *Sensors*, 23(9), 4258..

- [27] Zhou, X., Liang, W., Kevin, I., Wang, K., Wang, H., Yang, L. T., & Jin, Q. (2020). Deep-learning-enhanced human activity recognition for Internet of healthcare things. *IEEE Internet of Things Journal*, 7(7), 6429-6438..
- [28] Jegham, I., Khalifa, A. B., Alouani, I., & Mahjoub, M. A. (2020). Vision-based human action recognition: An overview and real world challenges. *Forensic Science International: Digital Investigation*, 32, 200901..
- [29] Berkeley MHAD: A Comprehensive Multimodal Human Action Database.
- [30] Kumar, P., Chauhan, S., & Awasthi, L. K. (2024). Human activity recognition (har) using deep learning: Review, methodologies, progress and future research directions. *Archives of Computational Methods in Engineering*, 31(1), 179-219..
- [31] Larsen, A. G., & Papi, G. (2023). Prediction of football actions and identification of optimal sensor placements using a semi-supervised learning approach..
- [32] Egidi, L., & Gabry, J. (2017). Bayesian hierarchical models for predicting individual performance in football (soccer). In *MATHSPORT INTERNATIONAL 2017 CONFERENCE-Proceedings*..
- [33] Noé, P. G. (2020). Emotion Recognition in Football Commentator Speech: Is the action intense or not?..
- [34] Salim, F. A., Postma, D. B., Haider, F., Luz, S., Beijnum, B. J. F. V., & Reidsma, D. (2024). Enhancing volleyball training: empowering athletes and coaches through advanced sensing and analysis. *Frontiers in Sports and Active Living*, 6, 1326807..
- [35] Huang, J., & Zou, W. (2023). Artificial Intelligence-based Volleyball Target Detection and Behavior Recognition Method. *International Journal of Advanced Computer Science and Applications*, 14(9)..
- [36] Meng, X., Xu, R., Chen, X., Zheng, L., Peng, A., Lu, H., ... & Zheng, H. (2018). Human action classification in basketball: a single inertial sensor based framework. In *Frontier Computing: Theory, Technologies and Applications (FC 2017)* 6 (pp. 152-161)..
- [37] Tang, B., & Guan, W. (2022). CNN Multi-Position Wearable Sensor Human Activity Recognition Used in Basketball Training. *Computational Intelligence and Neuroscience*, 2022(1), 9918143..
- [38] Mangiarotti, M., Ferrise, F., Graziosi, S., Tamburrino, F., & Bordegoni, M. (2019). A wearable device to detect in real-time bimanual gestures of basketball players during training sessions. *Journal of Computing and Information Science in Engineering*, 19.
- [39] Rosenhaim, A. L. (2023). Human Action Evaluation applied to Weightlifting..
- [40] Sensors | Free Full-Text | Automatic Swimming Activity Recognition and Lap Time Assessment Based on a Single IMU: A Deep Learning Approach (mdpi.com).

- [41] Brunner, G., Melnyk, D., Sigfússon, B., & Wattenhofer, R. (2019, September). Swimming style recognition and lap counting using a smartwatch and deep learning. In Proceedings of the 2019 ACM International Symposium on Wearable Computers (pp. 23-31)..
- [42] Radar-Based Swimming Activity Recognition with Temporal Dynamic Convolution and Spectral Data Augmentation (techrxiv.org).
- [43] Sensors | Free Full-Text | Detection of Tennis Activities with Wearable Sensors (mdpi.com).
- [44] Zhang, X., & Chen, J. (2023). A Tennis Training Action Analysis Model Based on Graph Convolutional Neural Network. IEEE Access..
- [45] Sensors | Free Full-Text | Temporal Pattern Attention for Multivariate Time Series of Tennis Strokes Classification (mdpi.com).
- [46] Jayakumar, B., & Govindarajan, N. (2024). Multi-sensor fusion based optimized deep convolutional neural network for boxing punch activity recognition. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Techno.
- [47] Vales-Alonso, J., González-Castaño, F. J., López-Matencio, P., & Gil-Castiñeira, F. (2023). A nonsupervised learning approach for automatic characterization of short-distance boxing training. IEEE Transactions on Systems, Man, and Cybernetics: Systems..
- [48] Baghel, V., Rithihas, N., Sarvanan, M., Srinivasan, B., & Hegde, R. S. (2024, May). Efficient boxing punch classification: fine-grained skeleton-based recognition made light., In Fifth International Conference on Computer Vision and Computational Intelligence (CVCI 2024) (Vol. 13169, p. 1316902). SPIE.
- [49] Mottaghi, A., Soryani, M., & Seifi, H. (2020). Action recognition in freestyle wrestling using silhouette-skeleton features. Engineering Science and Technology, an International Journal, 23(4), 921-930..
- [50] Gärtner, E., Pirinen, A., & Sminchisescu, C. (2020, April). Deep reinforcement learning for active human pose estimation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 10835-10844)..
- [51] Sozykin, K., Protasov, S., Khan, A., Hussain, R., & Lee, J. (2018, June). Multi-label class-imbalanced action recognition in hockey videos via 3D convolutional neural networks. In 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 146-151). IEEE.
- [52] Patel, S. H., Kamdar, D., Vyas, D. D., & Patel, P. P. (2023). DEEP LEARNING APPROACH FOR EVENT RECOGNITION IN FIELD HOCKEY VIDEOS. Reliability: Theory & Applications, 18(3 (74)), 316-332..

- [53] Faridee, A. Z. M., Ramamurthy, S. R., Hossain, H. S., & Roy, N. (2018, February). Happyfeet: Recognizing and assessing dance on the floor. In *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications* (pp. 49-54)..
- [54] Hendry, D., Chai, K., Campbell, A., Hopper, L., O'Sullivan, P., & Straker, L. (2020). Development of a human activity recognition system for ballet tasks. *Sports medicine-open*, 6, 1-10..
- [55] Host, K., & Ivašić-Kos, M. (2022). An overview of Human Action Recognition in sports based on Computer Vision. *Heliyon*, 8(6)..
- [56] Chen, X., & Yuan, G. (2021). Sports injury rehabilitation intervention algorithm based on visual analysis technology. *Mobile Information Systems*, 2021(1), 9993677..
- [57] LSTM with bio inspired algorithm for action recognition in sports videos - ScienceDirect.
- [58] Zhang, Y., & Hou, X. (2023). Application of video image processing in sports action recognition based on particle swarm optimization algorithm. *Preventive Medicine*, 173, 107592..
- [59] Radicchi, E., & Mozzachiodi, M. (2016). Social talent scouting: A new opportunity for the identification of football players?. *Physical Culture and Sport. Studies and Research*, 70(1), 28-43..
- [60] PANICHI, S. (2022). The selection of human resources in soccer: a weighted plus/minus metric for individual soccer player performance..
- [61] Guenter, R. W., Dunn, J. G., & Holt, N. L. (2019). Talent identification in youth ice hockey: Exploring "intangible" player characteristics. *The Sport Psychologist*, 33(4), 323-333..
- [62] "Wang, S. (2022). A Deep Learning Algorithm for Special Action Recognition of Football. *Mobile Information Systems*, 2022(1), 6315648."
- [63] "Wang, Q., Tao, B., Han, F., & Wei, W. (2021). Extraction and recognition method of basketball players' dynamic human actions based on deep learning. *Mobile Information Systems*, 2021(1), 4437146."
- [64] "Pan, Z., & Li, C. (2020). Robust basketball sports recognition by leveraging motion block estimation. *Signal Processing: Image Communication*, 83, 115784."
- [65] "Yao, B., Gao, H., & Su, X. (2020, November). Human motion recognition by three-view kinect sensors in virtual basketball training. In *2020 IEEE REGION 10 CONFERENCE (TENCON)* (pp. 1260-1265). IEEE."
- [66] "Gu, X., Xue, X., & Wang, F. (2020, May). Fine-grained action recognition on a novel basketball dataset. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2563-2567). IEEE."

- [67] "Jiang, L., & Zhang, D. (2023). Deep learning algorithm based wearable device for basketball stance recognition in basketball. *International Journal of Advanced Computer Science and Applications*, 14(3).".
- [68] "Francia, S. (2024). SpaceJam [Computer software]. GitHub. <https://github.com/simonfrancia/SpaceJam>".
- [69] "Hoelzemann, A., Romero, J. L., Bock, M., Laerhoven, K. V., & Lv, Q. (2023). Hang-time HAR: a benchmark dataset for basketball activity recognition using wrist-worn inertial sensors. *Sensors*, 23(13), 5879."
- [70] "Liu, R., Liu, Z., & Liu, S. (2021). Recognition of basketball player's shooting action based on the convolutional neural network. *Scientific Programming*, 2021(1), 3045418."
- [71] "Zhang, L. (2022). Behaviour detection and recognition of college basketball players based on multimodal sequence matching and deep neural networks. *Computational Intelligence and Neuroscience*, 2022(1), 7599685."
- [72] "Stoeve, M., Schuldhaus, D., Gamp, A., Zwick, C., & Eskofier, B. M. (2021). From the laboratory to the field: IMU-based shot and pass detection in football training and game scenarios using deep learning. *Sensors*, 21(9), 3071."
- [73] "Zhang, L. (2022). Applying deep learning-based human motion recognition system in sports competition. *Frontiers in Neurorobotics*, 16, 860981."
- [74] "Feng, C., & Wang, L. (2023). Analysis and Research on Technical and Tactical Action Recognition in Football Based on 3D Neural Network. *Applied Mathematics and Nonlinear Sciences*, 8(2), 1447-1462."
- [75] "Gao, Y., & Ma, G. (2021). Human motion recognition based on multimodal characteristics of learning quality in football scene. *Mathematical Problems in Engineering*, 2021(1), 7963616."
- [76] "Tsunoda, T., Komori, Y., Matsugu, M., & Harada, T. (2017). Football action recognition using hierarchical lstm. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 99-107)."
- [77] "Sultani, W., & Shah, M. (2021). Human action recognition in drone videos using a few aerial training examples. *Computer Vision and Image Understanding*, 206, 103186."
- [78] "Giancola, S., Cioppa, A., Georgieva, J., Billingham, J., Serner, A., Peek, K., ... & Van Droogenbroeck, M. (2023). Towards active learning for action spotting in association football videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and*
- [79] "Bose, S., Sarkar, S., & Chakrabarti, A. (2023, December). SoccerKDNet: A Knowledge Distillation Framework for Action Recognition in Soccer Videos. In *International Conference on Pattern Recognition and Machine Intelligence* (pp. 457-464). Cham: Springer Na".

- [80] "Raja, S., Kausalya, K., Sandhiya, B., Nihaal W, K., Mary, A., & Thahseen, J. A. (2024, January). Tracking of multi athlete and action recognition in soccer sports video using deep learning techniques. In AIP Conference Proceedings (Vol. 2802, No. 1). AIP".
- [81] S. C. A. G. J. B. J. S. A. P. K. .. & V. D. M. Giancola, "Towards Active Learning for Action Spotting in Association Football Videos," 2023.
- [82] "Giancola, S., Amine, M., Dghaily, T., & Ghanem, B. (2018). Soccernet: A scalable dataset for action spotting in soccer videos. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1711-1721).".
- [83] "Cioppa, A., Giancola, S., Somers, V., Magera, F., Zhou, X., Mkhallati, H., ... & Meng, Z. (2024). SoccerNet 2023 challenges results. Sports Engineering, 27(2), 24."
- [84] "Xu, J., Song, R., Wei, H., Guo, J., Zhou, Y., & Huang, X. (2021). A fast human action recognition network based on spatio-temporal features. Neurocomputing, 441, 350-358."
- [85] "Wang, L., Xu, Y., Cheng, J., Xia, H., Yin, J., & Wu, J. (2018). Human action recognition by learning spatio-temporal features with deep neural networks. IEEE access, 6, 17913-17922."
- [86] "Jaouedi, N., Boujnah, N., & Bouhlel, M. S. (2020). A new hybrid deep learning model for human action recognition. Journal of King Saud University-Computer and Information Sciences, 32(4), 447-453."
- [87] "Asghari-Esfeden, S., Sznaier, M., & Camps, O. (2020). Dynamic motion representation for human action recognition. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 557-566).".
- [88] "Yu, Z., & Yan, W. Q. (2020, November). Human action recognition using deep learning methods. In 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE."
- [89] "Kong, Y., Zhang, X., Wei, Q., Hu, W., & Jia, Y. (2008, December). Group action recognition in soccer videos. In 2008 19th International Conference on Pattern Recognition (pp. 1-4). IEEE."
- [90] "Barbon Junior, S., Pinto, A., Barroso, J. V., Caetano, F. G., Moura, F. A., Cunha, S. A., & Torres, R. D. S. (2022). Sport action mining: Dribbling recognition in soccer. Multimedia Tools and Applications, 81(3), 4341-4364."
- [91] "Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25."
- [92] "Namatēvs, I. (2017). Deep convolutional neural networks: Structure, feature extraction and training. Information Technology and Management Science, 20(1), 40-47."

- [93] "Vinyes Mora, S., & Knottenbelt, W. J. (2017). Deep learning for domain-specific action recognition in tennis. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 114-122).".
- [94] "Xu, S., Liang, L., & Ji, C. (2020). Gesture recognition for human-machine interaction in table tennis video based on deep semantic understanding. *Signal Processing: Image Communication*, 81, 115688.".
- [95] "Reno, V., Mosca, N., Marani, R., Nitti, M., D'Orazio, T., & Stella, E. (2018). Convolutional neural networks based ball detection in tennis games. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1758-1764).".
- [96] "Ó Conaire, C., Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., & Buckley, J. (2010, October). Combining inertial and visual sensing for human action recognition in tennis. In Proceedings of the first ACM international workshop on Analysis and retr".
- [97] "Zhu, G., Xu, C., Huang, Q., & Gao, W. (2006, August). Action recognition in broadcast tennis video. In 18th International Conference on Pattern Recognition (ICPR'06) (Vol. 1, pp. 251-254). IEEE.".
- [98] "Gourgari, S., Goudelis, G., Karpouzis, K., & Kollias, S. (2013). Thetis: Three dimensional tennis shots a human action dataset. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 676-681).".
- [99] "Liu, J., Wang, G., Hu, P., Duan, L. Y., & Kot, A. C. (2017). Global context-aware attention lstm networks for 3d action recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1647-1656).".
- [100] "Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, 31(7), 1235-1270.".
- [101] "Zhang, X., & Chen, J. (2023). A Tennis Training Action Analysis Model Based on Graph Convolutional Neural Network. *IEEE Access*.".
- [102] "Yan, S., Xiong, Y., & Lin, D. (2018, April). Spatial temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).".
- [103] Benages Pardo, L., Buldain Perez, D., & Orrite Urunuela, C. (2019). Detection of tennis activities with wearable sensors. *Sensors*, 19(22), 5004..
- [104] "Ji, X., & Liu, H. (2009). Advances in view-invariant human motion analysis: A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1), 13-24.".

- [105] "Buchanan, M. C. (1996, November). Technology-based community service. In Technology-Based Re-Engineering Engineering Education Proceedings of Frontiers in Education FIE'96 26th Annual Conference (Vol. 3, pp. 1352-1356). IEEE."
- [106] "Host, K., & Ivašić-Kos, M. (2022). An overview of Human Action Recognition in sports based on Computer Vision. *Heliyon*, 8(6)."
- [107] "Giancola, S., Cioppa, A., Georgieva, J., Billingham, J., Serner, A., Peek, K., ... & Van Droogenbroeck, M. (2023). Towards active learning for action spotting in association football videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and".
- [108] "Saiwa AI. (2024). Human pose estimation. Medium. <https://medium.com/@saiwa.ai/human-pose-estimation-349aed48973c>".
- [109] "Wang, Q., Tao, B., Han, F., & Wei, W. (2021). Extraction and recognition method of basketball players' dynamic human actions based on deep learning. *Mobile Information Systems*, 2021(1), 4437146."
- [110] Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2, 28..
- [111] Yao, B., Jiang, X., Khosla, A., Lin, A. L., Guibas, L., & Fei-Fei, L. (2011, November). Human action recognition by learning bases of action attributes and parts. In 2011 International conference on computer vision (pp. 1331-1338). IEEE..
- [112] Skublewska-Paszkowska, M., & Powroznik, P. (2023). Temporal pattern attention for multivariate time series of tennis strokes classification. *Sensors*, 23(5), 2422..