Kerbala Journal for Engineering Sciences

https://kjes.uokerbala.edu.iq

Advancements in Human Activity Recognition: A Comprehensive Survey of Sensor-Based and Vision-Based Approaches

Kawther Dhiya yassen *, Hawraa H.Abbas **, Arwa H. Mohammed Taqi***
* Department of Electrical and Electronic Engineering, University of Kerbala, Kerbala, Iraq
E-mail: kawther.d@s.uokerbala.edu.iq

**College of Information Technology Engineering, Al-Zahraa University for Women 56001 Karbala, Iraq
** Department of Electrical and Electronic Engineering, University of Kerbala Karbala 56001, Iraq
E-mail: Hawraa.h@alzahraa.edu.iq
*** Department of Electrical and Electronics Engineering, University of Kerbala, Kerbala, Iraq
E-mail: arwa.h@uokerbala.edu.iq

Received: 24 April 2024; Revised: 14 May 2024; Accepted: 25 September 2024

Abstract

The significance of human activity recognition (HAR) has increased in recent years because of its wide-ranging applications in areas such as healthcare, security and surveillance, entertainment, and intelligent settings. A significant challenge in computer vision is the automated and accurate recognition of human actions. This survey presents the last related work conducted over the 2015–2023 years in various areas of human activity recognition. It introduces the classification of HAR basic methodologies. In general, HAR approaches are categorized into two primary groups: sensor-based and vision-based HAR. This classification is based on the type of created data and the system environment itself. Based on our study, smart phones, smart watches, accelerometers, gyroscopes, and arm bands are all examples of sensor-based techniques. For the vision-based techniques, there are cameras, Microsoft Kinect, and thermal cameras. Also categorized the general datasets referenced in academic papers are based on the type of activity: individual actions, behaviors, interactions, and group activities. Subsequently, the preprocessing and feature engineering procedures are demonstrated. At long last, this review is able to offer some study concepts that investigate and analyze HAR.

Keywords: Human activity recognition (HAR), Vision-based, sensor-based, popular dataset.

1. Introduction

Researchers are becoming more and more interested in human activity recognition. Given the numerous potential applications of Human Activity Recognition (HAR), it is a highly researched field. HAR finds use in various scenarios, including medical applications, ambient assisted living, sports and leisure, tele-immersion, and security surveillance. Each of these diverse use cases has unique requirements, necessitating tailored approaches to meet specific needs.

For instance, a security surveillance system designed to detect criminal activities in public spaces inherently requires a vision-based approach, as it is impractical to equip individuals with sensors. Conversely, sports applications might involve wearable fitness trackers that automatically detect and monitor physical activities, or advanced fitness mirrors that recognize and count repetitions of specific body-weight exercises. As applications evolve, new use cases emerge, and novel devices are introduced, there is a growing need for continuously advancing HAR methodologies [1].

The development of human action recognition started in the early 1980s. Since then, research has predominantly concentrated on learning and identifying actions from video sequences captured by a single visible light camera. There is a substantial body of literature on action recognition across various disciplines, including computer vision, machine learning, pattern recognition, and signal processing [2].

2. Human Activity Recognition Approaches:

Activity recognition refers to the capacity to identify and detect ongoing activities by analyzing data obtained from various sensors. These sensors encompass various forms such as cameras, wearable sensors, sensors affixed to everyday objects, or deployed in the surrounding environment [3].

A variety of methods have been employed to document human behaviors. There are three main categories of approaches: vision-based, sensor-based, and hybrid-based [4]. as shown in Figure 1.

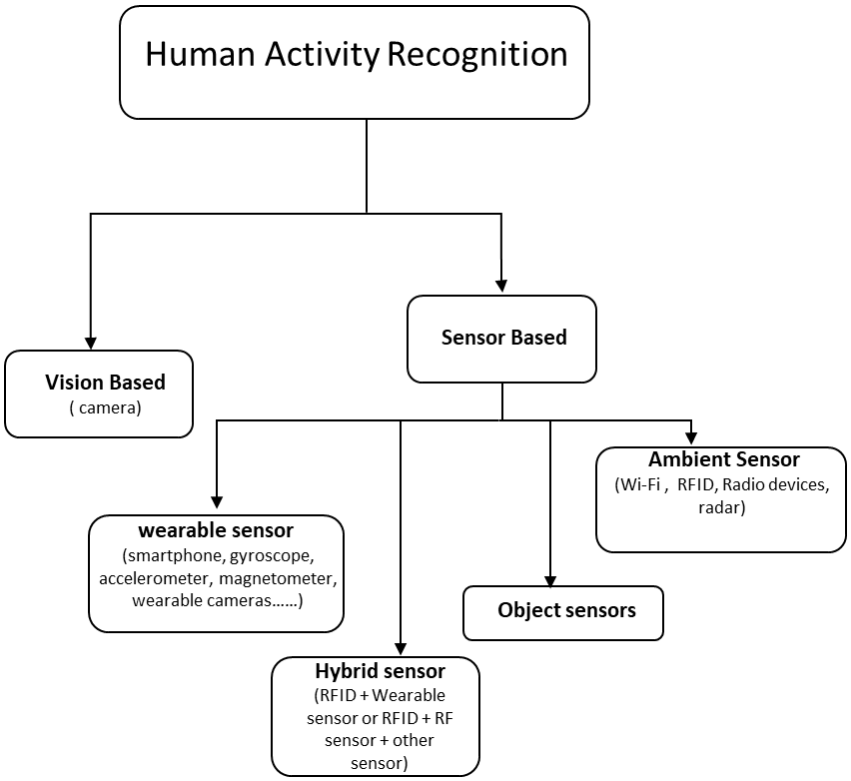

Figure 1:Classification of human activity recognition approaches.

vision-based approach

The vision-based human activity recognition (HAR) system employs cameras to monitor and analyze human actions and detect alterations in the surroundings. This methodology employs computer vision methodologies such as marker extraction, structural modeling, motion segmentation, action extraction, and motion tracking. Researchers employ a diverse range of cameras, including basic RGB cameras as well as more sophisticated systems that combine multiple cameras for stereo vision or utilize depth cameras capable of detecting scene depth using infrared lights [5].

Research on vision-based human activity recognition (HAR) can be categorized according to the type of data used, which comprises RGB data and RGB-D data. In general, vision-based human activity recognition (HAR) frameworks that utilize RGB data have demonstrated lower accuracy when compared to frameworks that utilize RGB-D data. This is because multi-modal data, such as RGB-D, provides

additional information and depth channels, which enhance the accuracy of the recognition process. Nevertheless, the extensive utilization of RGB data in current Human Activity Recognition (HAR) frameworks can be attributed to the presence of configuration complexity, computation complexity, and high costs [6].

- RGB data: A RGB image comprises of red, green, and blue bands within the visible spectrum, which can be captured by cameras equipped with a standard complementary metal-oxide-semiconductor (CMOS) sensor. RGB data is readily accessible, cost-effective, and yields detailed texture data of the subjects. Nevertheless, the sensor's range is restricted and can be affected by calibration issues. Additionally, it is highly dependent on environmental factors such as lighting, illumination, and the presence of a cluttered background [6].

Figure 2: RGB data.

- RGB-D data: RGB-D cameras capture depth data, which enhances the accuracy of algorithms in recognizing human activities. Each pixel in an RGB-D image represents the spatial distance between the screen space and an object in the RGB image. Pixels that are located near the camera have the highest values, while pixels that are far from the camera have the lowest values [6].

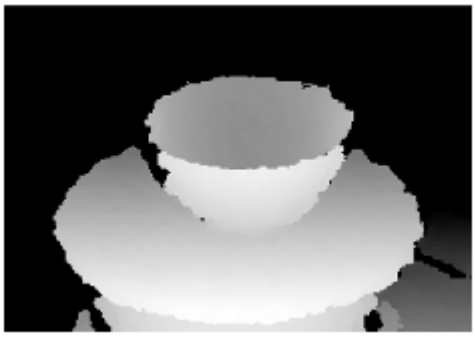

Figure 3: RGB-D data.

Sensor-based approach

The sensor-based human activity recognition (HAR) technique has been implemented in a range of practical applications, particularly in the domains of smart home technology and healthcare [6]. Sensor based involves the utilization of a network of sensors and connected devices to monitor and analyze an individual's physical movements and actions. The data they generate is presented as a time series of either state changes or parameter values. The diverse array of sensors, including touch detectors, RFID, accelerometers, motion sensors, noise sensors, and radar, can be directly attached to individuals, things, or the surrounding environment[5].

sensor-based approach divides into:

- Wearable sensors are the most common type of sensor used in HAR. Three common wearable sensors are the accelerometer, magnetometer, and gyroscope. These sensors can be easily worn by people or integrated into portable devices like smartphones, smartwatches, smart bands, glasses, or helmets. The detection of human actions can be achieved by detecting the variations in signals prior to and following the activity [6].
- Object sensors are sensors that are affixed to a specific object to detect and recognize activities associated with that object. Wearable sensors directly record human actions, whereas object sensors detect the movement of certain items to deduce human activity [6].
- Ambient Sensor (device free) the concept behind the Ambient Sensor approach is to place sensors in the surrounding environment (namely, the location where the activity is taking

place). These sensors will collect data whenever a person engages in any activity, which can then be utilized for activity identification purposes [3].

- Hybrid sensors are a fusion of various sensor types used in HAR applications, with the aim of enhancing the accuracy of activity recognition and the resilience of the model [6].

Sensor-based and vision-based approaches in human activity recognition vary in the types of data they use and the techniques they apply. Sensor-Based approaches Utilize data from 3-axis accelerometers and gyroscopes or others types of sensor to assess human activity, calculating movement speed, direction, and angle for precise recognition[7].

On the other hand, Vision-Based approaches, Analyze RGB images or videos with deep learning models such as ResNet, ViT or machina learning to identify human actions. While sensor-based methods emphasize processing sensor data to identify activity patterns and enhance recognition accuracy, vision-based techniques utilize visual data to model spatial interactions between entities. This enhances the understanding of human actions through image analysis and deep learning algorithms [8].

The primary challenges in vision-based human action recognition include intra-class variability and inter-class similarity of actions. Individuals may perform the same action in various directions and with different characteristics in body part movements, and some actions may be distinguishable only by subtle spatio-temporal details. Additionally, the vast number of describable action categories presents a challenge, as the same action may be interpreted differently depending on the context of objects and scenes. Furthermore, occlusions, cluttered backgrounds, cast shadows, varying illumination conditions, and viewpoint changes can all impact the perception of actions. The use of range cameras significantly mitigates the challenges associated with the third category, namely the common low-level difficulties that diminish recognition performance from 2D imagery. Additionally, range cameras offer valuable information about actions by capturing depth changes from specific viewpoints. For instance, in a frontal view, distinguishing between a person pointing and reaching is more accurate using depth map sequences compared to RGB footage [9].

Earlier range sensors faced several limitations: they were often prohibitively expensive, provided inaccurate estimations, or were challenging to use with human subjects. For instance, sonar sensors suffer from poor angular resolution and are prone to false echoes and reflections. Infrared and laser range finders can only measure a single point within a scene. LIDAR and radar systems are significantly more expensive and typically require higher power consumption. When using low-cost digital cameras, distance must be inferred through methods such as stereoscopic cameras or the motion of objects within the image, such as optical flow [9] .

3. Popular Datasets

The figure 4 illustrates the utilization of common datasets referenced in academic papers, classified based on the type of activity: atomic actions, behavior patterns, interactions, and group activities [10].

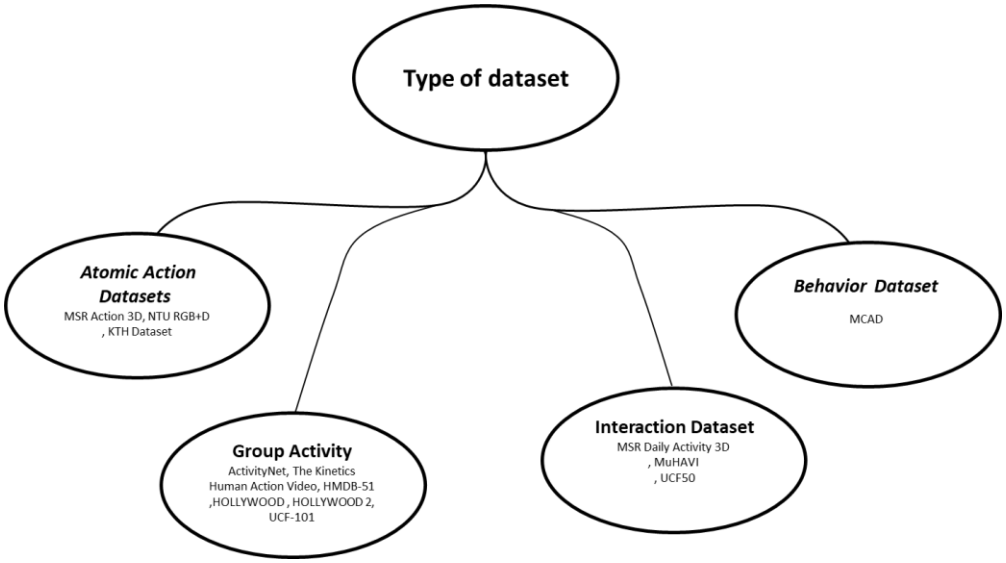

Figure 4: Classification of dataset.

- **Atomic Action Datasets**
  Atomic actions refer to basic movements performed at the atomic level, such as raising the hand or walking. They serve as the basis for more advanced deliberate and intentional actions [10].

  1. KTH Dataset
     The KTH dataset was generated by Sweden's Royal Institute of Technology in 2004. The dataset contains 2391 distinct actions across four different scenarios. There are a total of 25 distinct sets, each consisting of six different types of human action (running, jogging, walking, hand clapping, boxing, and waving). These activities are performed by 25 individuals and can be repeated up to five times. The videos have an average runtime of 4 seconds, consisting of chunks that are 4 seconds long. These segments are filmed on a stationary backdrop using a single camera [11].

Figure 5: Samples from KTH dataset.

2. NTU RGB+D
   The Nanyang Technological University generated this action recognition dataset in 2016. The HAR video library is vast, comprising more than 50,000 video segments and 4 million frames. The composition consists of 60 distinct performances, each executed by 40 distinct individuals, spanning tasks relating to health and social well-being. The dataset was acquired simultaneously using three Microsoft Kinect v2 sensors. The peculiarity of the object is attributed to its extensive range of viewing angles (80) from which it was observed. The expansion comprises a total of 120 distinct acts executed by 106 distinct individuals [12].

Figure 6: Sample from NTU RGB+D dataset

3. MSR Action 3D
   The MSR Action 3D was created at Microsoft Research Redmond under the supervision of Wanqing Li. There are 567 sets of depth maps in all. These maps capture the movements of 10 individuals performing 20 distinct actions, which are repeated either twice or three times. A Kinect gadget was utilized for capturing sequences [13].

Figure 7: Sample from MSR Action 3D dataset

- **Behavior dataset**
  External manifestations of individuals' emotional and psychological states that can be observed by others through their physical gestures and behaviors [10].
  1. Multi-Camera Action Dataset (MCAD)
     The dataset was developed by NUS-National University of Singapore in 2016. More precisely, its purpose was to evaluate the problem of categorizing open-view situations in a monitoring setting. Twenty participants were involved in recording eighteen

everyday activities from the KTH, TRECIVD, and IXMAS datasets using five cameras. Each activity was executed by a single person eight times for each camera, with four repetitions during the day and four repetitions during the night, in order to capture it [14].

Figure 8: Sample from MCAD dataset.

- **Interaction dataset**

These are the various manifestations of reciprocity in which alterations take place among the participants involved, whether they be individuals or objects [10].

1. MSR Daily Activity 3D Dataset

Jiang Wang, a researcher at Microsoft Research in Redmond, created this. The film consists of 320 depth maps, joint coordinates for each skeleton joint, and RGB clips of 10 individuals (both male and female) engaged in various activities, including eating, drinking, reading, and performing home duties [15].

Figure 9: Sample from MSR Daily Activity 3D Dataset.

2. Multi-Camera Human Action Video Dataset (MuHAVI)

The main focus is on strategies for recognizing human activities using silhouettes. The videos utilized have 14 artists executing their own action sequences on 14 occasions. The task was accomplished by employing eight unsynchronized cameras positioned on both the four sides and four corners of the platform [16].

Figure 10: Sample from MuHAVI dataset.

3. UCF50

The UCF50 was developed by the University of Central Florida's computer vision research institute in 2012. The theme for this project is that it is made up of 50 action classes, all taken from genuine YouTube videos. As an extension of the 11-category YouTube activity dataset (UCF11), this dataset features a wider variety of action-oriented videos [17].

Figure 11: Sample from UCF50 dataset

- **Group Activities dataset**

The collective actions carried out by multiple individuals, such as "embracing affectionately," are commonly known as "group activities." These actions may vary in complexity and can provide challenges in terms of monitoring and identification [10].

1. Activity Net Dataset

   This was implemented in 2015. The entire collection of 849 films is utilized to showcase more than 200 distinct activities, with each activity category containing 137 unedited videos. There exist three distinct algorithms for classifying human activity: unmodified video classification, activity classification without any filtering, and activity detection. The collection encompasses a wide range of complex human activities, comprising diverse circumstances and movements [18].

Figure 12: Sample from Activity Net Dataset.

2. The Kinetics Human Action Video Dataset
   This was built by the DeepMind team in 2017. The initial edition (Kinetics 400) consisted of 400 distinct categories of human activities, with each category containing a minimum of four hundred YouTube video samples showcasing a diverse array of activities. The Kinetics 600 dataset is an enhanced iteration of the previous Kinetics 400 collection, designed to encompass about 600 distinct human motion categories. Every action class in the Kinetics 600 dataset consists of at least 600 video clips. The Collection comprises over 500,000 brief movies, each with a duration of approximately 10 seconds and categorized under a single label [19].

Figure 13: Sample from Kinetics Human Action Video Dataset.

3. HMDB-51 Dataset
   The HMDB dataset, comprising over 7000 clips that have been tagged by hand and acquired manually from various sources such as YouTube, Google videos, and the Prelinger collection, was published by the Serre Laboratory at Brown University in 2011. The dataset for human action recognition consists of 51 classes, which are categorized into five distinct types of motion: human interaction, body movement, facial expression, object manipulation, and object interaction. The presence of ambient noise and unstable camera movements are two of the most challenging issues when working with authentic video material [20].

Figure 14: Sample from HMDB-51 Dataset.

4. HOLLYWOOD Dataset
   The dataset comprises diverse video clips and was initially introduced by the INRIA Institute in France in 2008. Each sample is labeled with one of eight actions: entering or exiting a vehicle, responding to a phone call, shaking hands, embracing, sitting, rising from a seated position, standing up, and kissing. The dataset was obtained from a total of 32 movies. Out of these, 20 movies were used to create a test set, while the remaining 12 movies were used to create the training sets [20].

5. HOLLYWOOD 2 Dataset
   INRIA released this in 2009 to augment the Hollywood dataset. The dataset has 12 distinct action categories, which are comparable to those found in the Hollywood dataset. However, it also contains four extra actions: driving a vehicle, getting in a car, eating, and fighting. In total, there are 3669 video clips, which were gathered from 69 movies and amount to nearly 20 hours of footage [21].

Figure 15: Sample from HOLLYWOOD 2 Dataset.

6. UCF-101 Action Recognition Dataset
   The UCF CRICV (Center for Research in Computer Vision) developed this in 2012 [17].The UCF101 dataset is an extension of the UCF50 dataset, which includes 50 different action types. A dataset including 13,320 videos from 101 real-world action classes was compiled by collecting videos from YouTube and merging them together. This offers the utmost flexibility in terms of movement and various viewpoints, as well as lighting conditions, among other factors [21].

4. Human Activity Recognition Framework

The human activity recognition framework consists of four primary components, as illustrated in Figure 15. The initial stage is the data acquisition phase, wherein data is obtained through optical sensing equipment or other types of sensors. The second phase is the pre-processing stage, which involves essential steps for processing the obtained data. The third stage involves the extraction of features from the dataset through methodologies such as machine learning and deep learning. The fourth phase involves the recognition or classification of activities[22].

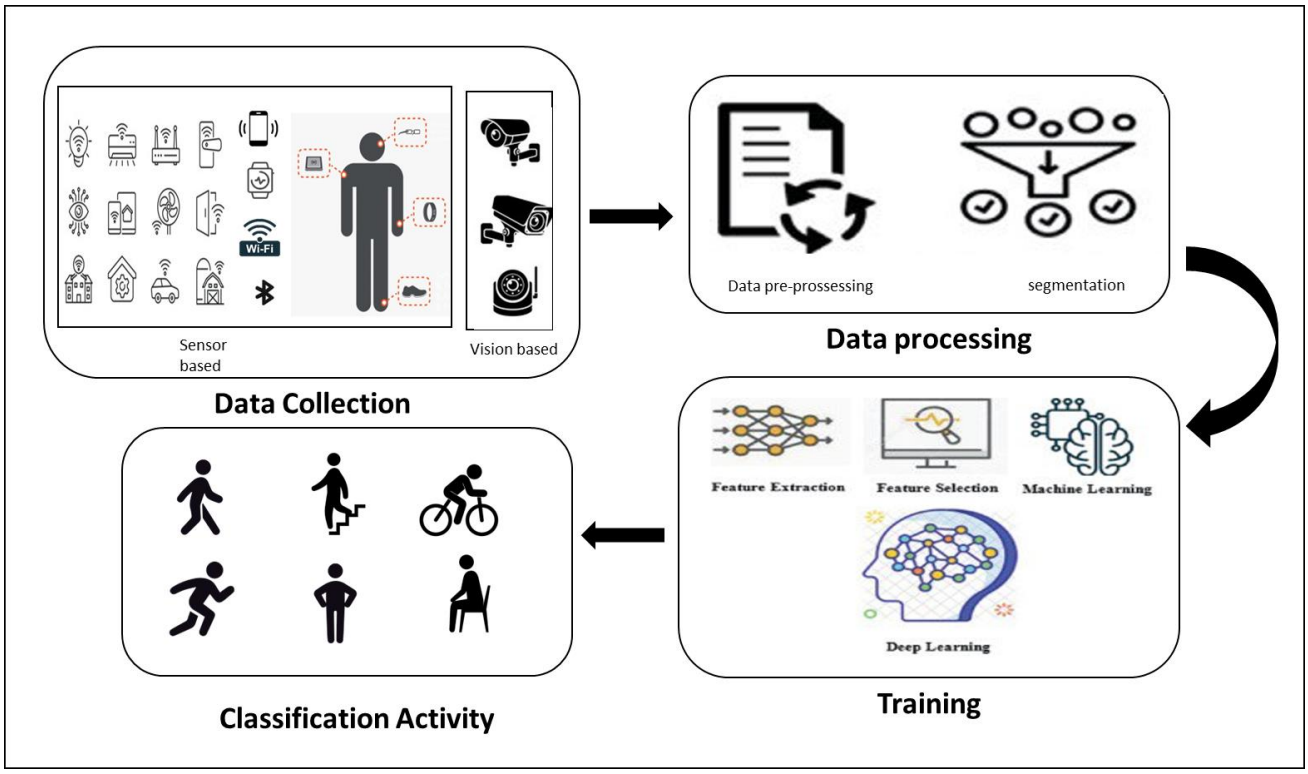

Figure 16: Human activity recognition framework.

5. Overview:

Depending on feature extraction and the types of activities being researched, research on human activity detection can be divided into several approaches. The human action identification techniques for accessible datasets can now be divided into two categories based on advancements in machine learning research: fully automated deep learning approaches or manually built features utilizing machine learning techniques. This section will examine various methods for human motion recognition.

In 2018 , Moeness G. Amin et al [23]. investigated the application of deep neural networks for radar-based human motion recognition, with an emphasis on the interpretability of classification results using the linear-wise relevance propagation (LRP) method. introduced the LRP method to identify relevant pixels in radar images, demonstrating its effectiveness in detecting regions essential for distinguishing human motions. The findings underscore the importance of understanding the impact of specific pixels on classification outcomes in deep neural networks applied to radar image analysis.

However, the experimental results presented demonstrate the successful application of the LRP method. Nonetheless, there is limited discussion on the potential challenges or drawbacks of using this method in real-world scenarios, such as its computational complexity or generalizability to diverse datasets.

Wearable sensors for human physiological monitoring have garnered significant interest from researchers; however, previous studies often lacked advanced analytical algorithms. Chi Cuong Vu et al. [24] combined textile stretch sensors based on single-walled carbon nanotubes (SWCNTs) and spandex fabric (PET/SP) with machine learning algorithms for human motion recognition. The study aimed to assess the system's performance based on identification rate and motion recognition accuracy, with the primary goal of developing a practical motion-sensing wearable product that eliminates the need for heavy and uncomfortable electronic devices. However, the research involved simple trials without significant analytical algorithms, limiting the depth of the analysis.

Fadhlan et al [25], Proposed a visual motion detection technique utilizing adaptive background subtraction with white pixel counts as a motion level identifier. Aims to develop a motion detection method that can accurately detect motion in non-static backgrounds, regardless of object speed, size, camera noise, or light intensity changes. The effectiveness of the proposed method in extremely low-light conditions or highly dynamic backgrounds is not extensively discussed and The paper does not explore the potential impact of varying camera qualities or resolutions on the performance of the proposed motion detection system

In 2017, Yuming Shao et al. [26], focused on radar-based human motion classification using high-resolution range information with a deep convolutional neural network (DCNN). Contrasting with existing methods that rely on handcrafted features from micro-Doppler signatures, this approach offers a more robust solution. The researchers collected real data from six human subjects performing seven motions, including walking, running, and boxing, to train and test the DCNN. The DCNN achieved an impressive accuracy rate of 95.24% in classifying human motions based solely on range information, showcasing its effectiveness. However, the small variation in range information mentioned in the paper may restrict the ability to generalize the findings to a broader range of scenarios or motions. Additionally, the study does not explore the real-time applicability of the proposed method, leaving room for further research on its practical implementation and efficiency in real-world scenarios.

In 2016, S. U. Park et al [27]. offer on Human Activity Recognition (HAR) in health and social care using a Depth Camera and Recurrent Neural Network (RNN). Previous studies have utilized classifiers such as the Hidden Markov Model (HMM) and Deep Belief Network (DBN) for HAR with depth human silhouettes. The proposed RNN-based HAR system outperforms HMM and DBN methods, achieving an average recognition accuracy of 99.55% for twelve activities on the MSRC-12 dataset. This

research emphasizes the importance of temporal changes in joint angle features for accurate activity recognition, highlighting the potential of RNNs in this domain.

In 2020, Meimei Gong et al. [28] introduced a new real-time detection and recognition network called IMFF-SSD, based on the Single Shot MultiBox Detector (SSD). This network aims to enhance the accuracy of detecting and recognizing human moving targets in videos. IMFF-SSD combines low-level detail features with high-level semantic features through a process called multi-scale feature fusion, which improves the positioning accuracy of the target prediction layer in the SSD network. The authors designed a feature-embedded prediction structure that strengthens the semantics of target features without altering the spatial resolution, thereby enhancing the accuracy of recognizing human moving targets at various scales. Experimental results demonstrate that the IMFF-SSD network significantly improves both positioning accuracy and recognition accuracy compared to the original SSD network, making it more effective for real-time human motion detection and recognition. However, the researchers do not address the computational complexity and resource requirements of the proposed IMFF-SSD network, which could be a limitation for real-time applications on devices with limited processing power. Additionally, the study lacks a detailed analysis of the network's performance under varying lighting conditions or occlusions, which are common challenges in real-world scenarios.

Fuquan Zhang et al. [29] in 2019 proposed a new method for recognizing human motion in virtual reality (VR) environments by combining linear discriminant analysis (LDA) with a support vector machine (SVM) algorithm, effectively addressing nonlinear problems and expanding sample differences. The authors utilized a kernel function in the LDA algorithm to extract high-dimensional nonlinear features from motion data, crucial for accurate motion recognition in complex VR scenarios. A genetic algorithm was employed to optimize the parameters of the SVM classifier, leveraging its strengths in multi-dimensional space optimization to enhance the accuracy and robustness of motion classification.

The study compared the proposed LDA-GA-SVM method with traditional K-means-SVM algorithms, demonstrating that the new method achieves higher accuracy, specificity, and sensitivity in human motion recognition tasks. However, due to the complexity and diversity of human motion in VR scenes, some critical high-dimensional nonlinear feature information hidden in motion data may not be effectively extracted, potentially affecting the overall performance of the recognition system. Additionally, the proposed method relies heavily on the genetic algorithm for parameter optimization, which, while effective, can be computationally intensive and time-consuming, potentially limiting its scalability for real-time applications

In 2020 Danica Hendry et al. [30] developed a system using wearable sensors and machine learning to accurately identify key ballet movements, such as jumping and leg lifts, during dance training. The study demonstrated that convolutional neural networks could achieve high accuracy, with the best model performing at 97.8% accuracy when using data from all sensors and excluding transition movements. The research showed that the accuracy of the models decreased when fewer sensors were used, when transition movements were included, and with different sensor combinations, highlighting the importance of sensor placement and quantity.

The system provides a new way to measure dancer training volume, which can help understand the relationship between dancers' pain and their training volume. This approach can also be applied to other sports that focus on lower limb movements. However, the accuracy of the models decreased when fewer sensors were used, when transition movements were included, and with different sensor combinations, indicating that the system's performance is highly dependent on the number and placement of sensors. Additionally, the study focused only on specific ballet movements like jumping and lifting the leg, which means the system may not be applicable to other types of dance movements or activities without further development and validation.

In 2022, Guilherme Augusto Silva Surek et al. [31] focused on evaluating and mapping the current state of human activity recognition in videos using deep learning models, specifically ResNet and Vision Transformer (ViT) architectures, to better understand human actions. The study introduced a semi-supervised learning approach called DINO (self-DIstillation with NO labels) to enhance the performance of ResNet and ViT models, improving the accuracy of recognizing human activities in videos. The research utilized the HMDB51 dataset, a benchmark for human motion, to test the models, aiming to capture the richness and complexity of human actions more effectively. The results demonstrated that the proposed ViT model, combined with long short-term memory (LSTM), achieved high performance in human action recognition, with 96.7% accuracy in the training phase and 41.0% accuracy in the testing phase on the HMDB51 dataset. This discrepancy indicates potential overfitting issues, where the model performs well on training data but poorly on unseen data. Additionally, while the use of a semi-supervised learning approach like DINO is innovative, it may still require substantial labeled data to achieve optimal performance, which can be a limitation in scenarios where labeled data is scarce.

In 2023, Jaegyun Park et al [32]. conducted a review of several pedestrian datasets, including INRIA, Caltech, MS COCO, KITTI, and ETH, noting their limitations such as small size, limited scenarios, and challenges with illumination and occlusion. The researchers examined advanced deep learning models like Mask R-CNN, R-FCN, and RetinaNet, which have been effective in pedestrian detection and tracking, especially under difficult conditions like varying lighting and occlusions. This study compared the performance of these models on well-known datasets such as PASCAL VOC and MS COCO, showcasing their proficiency in tasks like pedestrian identification and suspicious activity recognition. Furthermore, a new dataset was introduced, which captures student behavior in academic settings and provides detailed annotations and stable ID tracking for pedestrians, making it suitable for real-time surveillance applications. Despite these advancements, the improved Mask R-CNN model still encounters difficulties with varying illumination and occlusion, impacting the accuracy of pedestrian detection and tracking. Additionally, the computational complexity of these advanced deep learning models, such as the Improved Mask R-CNN, may limit their real-time performance, making them less practical for immediate surveillance needs.

Keyou Guo et al [33]. introduced a new neural network model called NEW-STGCN-CA, designed for human action recognition by focusing on both local and global information in skeleton data. The model incorporates a Coordination Attention (CA) mechanism, which helps the network focus on important input-related information while ignoring unnecessary details, thereby preventing information loss and improving accuracy. A new partitioning strategy is proposed to enhance the connection between local and global information, significantly improving the model's feature extraction capabilities and

robustness. Experimental results show that the NEW-STGCN-CA model achieves higher accuracy compared to the original ST-GCN model on the NTU-RGBD 60 and Kinetics-Skeleton datasets, demonstrating its effectiveness and robustness. However, the authors do not address the computational complexity and resource requirements of the NEW-STGCN-CA model, which could be a concern for real-time applications and deployment on devices with limited processing power. Additionally, while the new partitioning strategy improves accuracy, the research does not provide a detailed analysis of its impact on the model's training time and convergence rate, which are important factors for practical implementation.

Most studies in Human Activity Recognition (HAR) have focused on improving accuracy through methods such as integrating short-time Fourier transform into preprocessing steps and using 2D CNNs. However, these often require complex preprocessing, increasing overhead. Ravi et al. (2016) and Bhat et al. [34] (2018) attempted real-time HAR using techniques like discrete wavelet transform and deep neural networks, but these approaches also involve cumbersome preprocessing steps. Recent studies have shifted towards using convolutional neural networks (CNNs) for HAR due to their lower computational costs and suitability for real-time responses. Nonetheless, manually designing CNNs may not always achieve optimal accuracy due to varying computational budgets. The proposed GTSNet framework in this paper aims to address these issues by using a mathematical approach to derive a CNN architecture that can be flexibly modified to match desired computational budgets, demonstrating superior performance on benchmark datasets under limited computational resources. However, the GTSNet framework does not perform as well on more complex datasets like OPPORTUNITY and PAMAP2, indicating that it may require more sophisticated network architectures to classify complex activities correctly. Achieving real-time response in real-world HAR applications remains impractical due to the challenge of balancing computational efficiency and accuracy with the current GTSNet design.

Hakan Bilen et al. [35] introduced the concept of the 'dynamic image,' a novel way to represent videos in a compact form by summarizing video dynamics and appearance into a single image using a method called 'rank pooling.' This approach demonstrates that dynamic images can be used with existing convolutional neural networks (CNNs) pre-trained for still images, facilitating the extension of these models to video analysis without the need to train new models from scratch. The authors present an efficient and effective approximate rank pooling operator, significantly speeding up standard rank pooling algorithms and making the process more practical for real-world applications. They also formulate this approximate rank pooling operator as a CNN layer, allowing the generalization of dynamic images to dynamic feature maps, which can be utilized for more complex video analysis tasks. However, the method's reliance on summarizing video dynamics into a single image might result in the loss of some fine-grained temporal information, which could be crucial for recognizing subtle actions or events in videos.

Munkhjargal Gochoo et al. [36] in 2017 introduced a novel method for recognizing the activities of elderly people living alone using a Deep Convolutional Neural Network (DCNN) and binary sensors, which do not invade privacy. For the first time, the researchers converted sequences of data from PIR sensors into activity images, which are then used for deep learning to improve activity recognition. The study utilized an open dataset collected over two years from the Aruba testbed, provided by the Center for Advanced Studies in Adaptive Systems (CASAS) project, to validate the proposed DCNN classifier.

The proposed DCNN classifier achieved the highest performance in activity recognition among systems using binary sensors, with an accuracy of 99.36%. However, the authors focused on recognizing only four basic activities (Bed_to_Toilet, Eating, Meal_Preparation, and Relax), which may not cover the full range of daily activities that elderly people perform, limiting its applicability in real-world scenarios. While the classifier achieved high accuracy, the study does not provide detailed information on the computational cost and real-time performance, which are important for practical implementation in smart homes.

In 2023, Ahmed M. Helmi et al. [37] use wearable sensor data to combine the applications of swarm intelligence (SI) and deep learning (DL) to create a reliable HAR system. employing a recurrent neural network in conjunction with a residual convolutional network (RCNN-BiGRU). Create techniques based on the marine predator algorithm (MPA) to choose the best feature set. and making use of the UniMiB-SHAR, PAMAP2, and The Opportunity databases.

In 2022, Vidhi Jain et al. [38] introduced a new method for recognizing human actions using a combination of Bi-Convolutional Recurrent Neural Network (Bi-CRNN) for feature extraction and Random Forest for classification, enhancing the accuracy and efficiency of human action recognition in autonomous systems. This approach leverages ambient intelligence, utilizing data from various sensors in the environment to understand human actions, making it highly relevant for applications in smart environments and autonomous robots. The method includes an auto-fusion technique that improves the integration and processing of data from multiple sensors, resulting in more accurate human action recognition compared to existing algorithms. The results show a high accuracy rate of 94.7% for human action recognition, demonstrating the effectiveness of the hybrid deep learning-based algorithm in practical applications. However, the study focuses on ambient intelligence for autonomous robots, which may limit its applicability to other domains such as healthcare or smart homes, where different types of sensors and data might be used.

In 2021, Muhammad Muaaz et al. [39] introduced Wi-Sense, a human activity recognition (HAR) system that utilizes Wi-Fi signals and a convolutional neural network (CNN) to identify human activities, aiming to remain independent of environmental changes. Wi-Sense captures Wi-Fi channel state information (CSI) and applies methods such as the CSI ratio method and principal component analysis to reduce noise, eliminate redundant information, and minimize environmental impact. The system was tested with data from nine volunteers in an indoor environment, achieving an overall accuracy of 97.78% in recognizing various activities. Although Wi-Sense aims to be environment-independent, the system's performance can still be affected by significant changes in the ambient environment, which may introduce noise and impact accuracy. Additionally, the process of capturing and processing Wi-Fi channel state information (CSI) involves multiple steps, such as noise reduction and principal component analysis, which can be computationally intensive and may require significant processing power.

Md. Milon Islam et al. [40] introduce a novel approach that integrates various data types using deep learning techniques to enhance human activity recognition, particularly for smart healthcare applications. The research employs Convolutional Neural Networks (CNNs) to extract key features from images and Convolutional Long Short Term Memory (ConvLSTM) to identify patterns in multi-sensory data, improving the overall activity recognition process. A self-attention mechanism is used to focus on relevant activity data while disregarding unnecessary information, thus boosting the accuracy and

reliability of the recognition system. This method has been rigorously tested with the UP-Fall detection dataset, showing superior performance compared to existing state-of-the-art techniques in terms of efficiency and robustness. However, the use of CNNs and ConvLSTM models can be computationally intensive, requiring substantial processing power and memory, which may not be feasible for all healthcare settings.

Device-free human Activity Recognition and Monitoring System (DARMS), presented by Zhihao Gu et al. in 2022 [41] can be implemented using inexpensive commodity WiFi devices. which is a hardware module for data gathering, a software module for signal preprocessing, and a neural network for activity recognition make up this passive wireless sensing system's three main parts. This method makes use of the CSI dataset and convolutional neural networks. DARMS performs exceptionally well in a variety of interior settings, with 96.9% accuracy in fall detection and 93.3% accuracy in human activity recognition.

Muhammad Atif Hanif et al. [42] introduce a new framework capable of recognizing both basic and complex human activities using the built-in sensors of smart devices like smartphones and smartwatches. The framework considers 25 different physical activities, including 20 complex ones, which is more extensive than existing studies that typically consider up to 15 activities. By utilizing the sensors already present in widely available smart devices, the framework avoids the need for additional external devices, making it cost-effective and convenient. One of the main limitations is the lack of labeled datasets for complex human activities, which makes it challenging to train and validate the proposed framework effectively. Additionally, the framework relies heavily on the built-in sensors of smart devices, which may not be uniformly available or consistent across different brands and models.

In 2022, Sarah Khater et al [43] Introduce a new layer in this study called residual inception convolutional recurrent layer (ResIncConvLSTM), which is a modification of the ConvLSTM layer . This is tested against the KTH and Weizmann datasets in addition to being trained on the KTH dataset. Additionally, the architectures are trained and tested against a portion of the UCF Sports Action dataset. For the ResIncConvLSTM architecture, the architectures perform well in recognizing boxing, clapping, and waving activities with accuracy rates of 99.9%, 99.5%, and 99.7%, respectively.

A system for real-time activity recognition is proposed by Raúl Gómez Ramos et al. in 2022 [44]. This system is based on various Deep Learning (DL) techniques that use neural networks, including Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), and Gated Recurrent Unit networks (GRU). The primary goal is to build a database by utilizing three distinct technologies in tandem to gather data on the everyday routines of numerous individuals residing in the same home (SDHAR-HOME). The prediction system's 90.91% accuracy .

Muhammad Atif Hanif et al. [45] introduce a new framework capable of recognizing both basic and complex human activities using the built-in sensors of smart devices like smartphones and smartwatches. The framework considers 25 different physical activities, including 20 complex ones, which is more extensive than existing studies that typically consider up to 15 activities. By utilizing the sensors already present in widely available smart devices, the framework avoids the need for additional external devices, making it cost-effective and convenient. One of the main limitations is the lack of labeled datasets for complex human activities, which makes it challenging to train and validate the

proposed framework effectively. Additionally, the framework relies heavily on the built-in sensors of smart devices, which may not be uniformly available or consistent across different brands and models.

In 2021, Raúl Gómez Ramos et al. [46] introduced a system utilizing Artificial Intelligence Technologies (AIT) to recognize the daily activities of elderly individuals in real-time, assisting specialists in monitoring habits such as taking medication or eating meals, thereby enhancing their quality of life and safety at home. The system employs a prediction model based on bidirectional Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, to accurately identify activities using data from various sensors installed in the homes of elderly individuals. The model achieves a high accuracy rate of 95.42%, which surpasses that of similar models currently in use, making it a reliable tool for real-time activity recognition. However, the system heavily depends on the accuracy and placement of the sensors in the home, meaning any misplacement or malfunction of sensors could result in incorrect activity recognition, affecting the overall reliability of the system.

Driver action recognition model was proposed by Mingqi Lu et al in 2019 [47] which is called deformable and dilated Faster R-CNN (DD-RCNN). In this model, the improved ResNet incorporates attention modules to reweight features in both the channel and spatial dimensions. The region proposal optimization network (RPON) is then introduced to increase model efficiency and decrease the amount of ROIs entering the R-CNN. Results of experiments using SEU-real-driving and Kaggle-driven datasets. accuracy rate 92.1% for SEU-real-driving and 86.0% for Kaggle-driving dataset.

In 2021, Thi Thi Zin et al [48] presented a vision-based system that uses depth maps captured by a Stereo depth camera to monitor and identify the varied movements of old individuals. An SVM classifier uses the feature vector to identify the action. Using a dataset gathered from an elder care center, the experiment tested and assessed the system. 95.6% accuracy.

In 2023, Roberta Vrskova et al [49] conducted research on video categorization using a combination of 3DCNN and ConvLSTM networks. While 3DCNN networks make use of the third dimension for classification, ConvLSTM networks use their temporal memory to identify spatiotemporal patterns in videos. The precision of the entire set of experimental results on the LoDVP Abnormal Activities dataset, UCF50 dataset, and MOD20 dataset was 89.12% when using the LoDVP Abnormal Activities dataset, and 83.89% and 87.76%, respectively, when using the modified UCF50 dataset (UCF50mini) and MOD20 dataset.

A deeply coupled ConvNet, which uses RGB frames at the top layer with bi-directional long short-term memory (Bi-LSTM), was proposed by Tej Singh [50] in 2020 for the recognition of human movement. One dynamic motion picture is used to train the CNN model at the lowest layer. Four common single- and multiple-person activities are used to evaluate the model's performance accuracy (The SBU Interaction, MIVIA Action, MSR Daily Activity 3D, MSR Action Pairs 3D dataset). for the MSR Daily Activity Dataset the accuracy was 94.37%, for the MIVIA Action Dataset the accuracy was 99.4%,for the SBU Interaction Dataset the accuracy was 98.7% and for the MSR Action Pairs Dataset the accuracy was 98.3%.

In 2019, Zhelong Wang et al [51] presented the SwimSense monitoring system, which uses wearable inertial sensors to track swimmers' progress. The Hidden Markov Model (HMM) is the classifier that is employed. This model recognizes four distinct swimming strokes using three different types of sensor data: magnetometer, gyroscope, and acceleration data. The results of the recognition process with gyroscope and magnetometer data are the worst, with an accuracy of 86:83%. The

recognition accuracy with acceleration and gyroscope data is 94:38% and with magnetometer and acceleration data is 91:84%

In 2021, Amin Ullah et al [52] suggests a framework for activity recognition with lightweight deep learning-assisted. which use an efficient CNN model trained on two surveillance datasets to identify a human in the surveillance stream. Using an extremely quick object tracker known as the "minimum output sum of squared error," the identified person is followed throughout the video stream (MOSSE). Next, using the effective LiteFlowNet CNN, pyramidal convolutional features are recovered from two consecutive frames for each tracked individual. Lastly, a novel deep skip connection gated recurrent unit (DS-GRU) is trained to recognize activities by learning the temporal variations in the frame sequence. F1-scores for this model were 82.12%, 88.63%, 60.73%, and 65.79%. and 91.43% for the Hollywood2, UCF-101, UCF-50, HMDB51, and datasets for YouTube Actions, correspondingly

In 2021, Rajesh Amerineni et al [53] presented a collection of models that combine data from wearable sensors to identify general human movement classes without relying on any evaluated movement characteristics. Convolution neural network (CNN) and dynamic time warping (DTW) classifiers were chosen for the creation of movement classification systems. The 18-class boxing data, with 95% accuracy.

The approaches suggested by TAMER SHANABLEH [54] in 2023 are predicated on ideas related to video coding, such as motion compensations and feature variables based on coding. These characteristics are employed in deep learning to create and classify models. A video input is momentarily divided into twelve equal-sized, non-overlapping segments, each of which is then transformed into a single RGB image component. An LSTM is used for classification and a Convolutional Neural Network (CNN) network is used for training. accuracy for the jHMDB dataset was 78.5%, the HMDB51 dataset had 71.4%, and the UCF11 dataset had 97%.

The SlowFast approach that Gyu-Il Kim et al [55] suggested in 2023, The suggested technique enhances SlowFast's data structure and preprocessing techniques for input data to obtain a high degree of extraction and accuracy. YOLO and DeepSORT are used for object tracking and background removal in the preprocessing of the incoming data. Accuracy values of 70.16% and 70.74% were attained for the suggested model and the current SlowFast, respectively.

Shangbin Li et al. [56] present a method for recognizing human motion using Nano-CMOS image sensors, enhancing the transformation and processing of human motion images. This technique integrates the background mixed model of pixels in human motion images to extract and select pertinent features, thereby increasing feature extraction accuracy. It exploits the three-dimensional scanning capabilities of Nano-CMOS sensors to gather detailed human joint coordinate data and sense motion state variables, resulting in a more precise human motion model. By calculating feature parameters for each motion gesture, the method captures foreground features of human motion images, improving both recognition accuracy and speed. The approach achieves a notable human motion recognition rate of 92% and a recognition speed of up to 186 frames per second, showing significant advancements over traditional methods. However, the authors do not address the potential challenges and limitations of using Nano-CMOS image sensors in various environmental conditions, such as low light or high motion scenarios, which could impact the accuracy of human motion recognition.

A human motion recognition technique based on passive RFID and multi-model fusion is proposed by Xu Yang et al [57]in 2023. by affixing a liquid metal tag that is bendable to the human body. The Blending model is created by fusing the four models of KNN, DT, SVM, and LR. For motion, LR serves as a secondary learner while KNN, DT, and SVM serve as primary learners. The results of the studies demonstrate that the Blending model has a 97.29% accuracy rate for the five actions of standing, sitting, walking, running, and falling.

In 2017, Mariofanna Milanova et al. [58] introduced a new model that employs a 3D deep neural network (3D DNN) to recognize human actions from video frames, tested under various conditions from surveillance cameras. This model is trained using the CaffeGoogLeNet framework with different training epoch values (TEs) and evaluated on three datasets: KTH, Weizmann, and UCF101, achieving high classification accuracy and short running times. The method consists of two main steps: extracting features from video frames using convolution and pooling layers, and classifying human actions using fully connected and softmax classifiers. Utilizing high parallelism multi-GPU hardware significantly reduces the running time, enhancing the model's performance and making it suitable for real-time applications. However, the running time of the model increases linearly with the number of training epochs, which could be a limitation for applications requiring real-time processing.

Arwa Mohammed Taqi et al. [59] introduced a new model called HMIV (Hu Moment Invariants on Videos) for recognizing human actions in videos. This model is robust to changes in scale, rotation, and translation, making it effective under various conditions such as different sides, positions, directions, and lighting. The authors use Hu moments to extract features from human action video sequences, averaging these moments across all frames to create a dominant feature for each video clip, thereby improving recognition accuracy. The method was tested on two datasets, KTH and UCF101, achieving high classification accuracies of 93.4% and 92.11%, respectively, demonstrating its effectiveness compared to other state-of-the-art techniques. The research also compared the performance of the HMIV method with other existing methods, showing that HMIV provides better results in recognizing human actions from video clips. However, the classification process uses Euclidean distance, which may not be the most sophisticated or accurate method for distinguishing between very similar actions, potentially affecting overall accuracy in more challenging datasets.

6. Conclusions

Efficiently comprehending and interpreting human activities is necessary in various fields of computer vision, such as human-computer interaction, robotics, monitoring, and security.

This study initially explored the many approaches to human activity recognition, which may be categorized into two groups: vision-based systems and sensor-based systems. The vision-based Human Activity Recognition (HAR) system uses cameras to observe and analyze human behavior, detecting changes in the environment. Another type of HAR system involves a network of sensors and interconnected devices to monitor and record a person's activities. Sensor-based systems can be categorized into three different categories: wearable sensors, sensors placed on objects, ambient sensors, and hybrid sensors. As a result of the reduced cost and improved sensor technology, the majority of research in the field of human activity recognition (HAR) has transitioned to a method that utilizes a sensor-based approach. However, wearable sensors also encounter numerous issues, with the primary one being the impracticality of wearing a tag in certain situations. For instance, when it comes to elderly individuals or patients, they may inadvertently neglect to wear the tags, or perhaps they actively refuse to wear them altogether. Additionally, relying solely on object sensors may not always be practical since it restricts users to using tagged objects. However, the ambient sensor is more efficient as it eliminates the need for the user to carry any device during any activity. However, there are drawbacks to this strategy, such as potential environmental interference. The surrounding environment has the potential to

interfere with the sensors' data collection, causing noise in the data. Subsequently, it presented an overview of the widely used datasets cited in scholarly articles and classified them according to the nature of the activity: atomic actions, behavioral patterns, social interactions, and group activities. Furthermore, the human activity recognition framework involves four fundamental components: data collection, data processing, training, and classification. Finally, a thorough summary of the studies conducted on human activity recognition was provided in table 1.

Table 1: A summary of the reviewed researches for human activity recognition.

NO	Research Title	Name of Researchers and Years	Algorithm	Dataset	Result accuracy
1	Understanding Deep Neural Networks Performance for Radar-based Human Motion Recognition	Moeness G. Amin and Baris Erol 2018	linear-wise relevance propagation (LRP)	radar images of human motion	89.2%.
2	Human Motion Recognition by Textile Sensors Based on Machine Learning Algorithms	Chi Cuong Vu and Jooyong Kim 2018	textile stretch sensors and machine learning algorithms (random forest (RD), support vector machine (SVM), one-hidden layer neural network (ANN), multi-hidden layers neural network (MANN), and autoencoders neural network)	Real motion dataset (walking, running, sprinting, jumping)	90% (RD), 84% (SVM), 85% (ANN), 88% (MANN)
3	Efficient Human Motion Detection with Adaptive Background for Vision-Based Security System	Fadhlan Hafizhelmi Kamaru Zaman, Md. Hazrat Ali, A.A. Shafie 2017	Adaptive background subtraction	real-time video streams	99.1%
4	Human Motion Classification Based on Range Information with Deep Convolutional Neural Network	Yuming Shao, Sai Guo, Lin Sun, Weidong Chen 2017	Deep Convolutional Neural Network (DCNN)	Real data collected using an ultra-wideband (UWB) radar	95.24%
5	A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services	S. U. Park, J. H. Park, M. A. Al-masni, M. A. Al-antari, Md. Z. Uddin, T. -S. Kim 2016	Recurrent Neural Network (RNN) based Human Activity Recognition (HAR)	MSRC-12 dataset	99.5%
6	Real-time Detection and Motion Recognition of Human Moving Objects Based on Deep Learning and Multi-scale Feature Fusion in Video	Meimei Gong, Yiming Shu 2020	multi-scale feature fusion (IMFF-SSD)network	video	96.8%
7	Human motion recognition based on SVM in VR art media interaction environment	Fuquan Zhang, Tsu-Yang Wu ,Jeng-Shyang Pan, Gangyi Ding and Zuoyong Li 2019	Linear discriminant analysis , genetic algorithm and support vector machine LDA-GA-SVM	Motion capture of the human body using inertia motion capture	94.97%
8	Development of a Human Activity Recognition System for Ballet Tasks	Danica Hendry, Kevin Chai, Amity Campbell, Luke Hopper, Peter'Sullivan1 and Leon Straker 2020	Convolutional neural networks (CNN) with wearable sensor	Real data collection from ballet studio	97.8%
9	Video-Based Human Activity Recognition Using Deep Learning Approaches	Guilherme Augusto Silva Surek, Laio Oriel	residual network (ResNet) and a vision	Database(HMDB51)	41.0 _ 0.27%

		Seman, Stefano Frizzo Stefenon, Viviana Cocco Mariani and Leandro dos Santos Coelho 2022	transformer architecture (ViT) with a semi-supervised learning)		
10	Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Activity Recognition	Ujwalla Gawande, Kamal Hajari, Yogesh Golhar 2023	YOLOv5 detector and Mask R-CNN	Microsoft COCO pedestrian dataset	83.10%
11	A New Partitioned Spatial–Temporal Graph Attention Convolution Network for Human Motion Recognition	Keyou Guo, Pengshuo Wang , Peipeng Shi, Chengbo He and Caili Wei 2023	spatial–temporal graph convolution network (ST-GCN),	NTU-RGB+D 60 And Kinetics-Skeleton dataset	84.86% for NTU-RGB+D 60 dataset , 32.40%. for Kinetics-Skeleton dataset
12	GTSNet: Flexible architecture under budget constraint for real-time human activity recognition from wearable sensor	Jaegyun Park, Won-Seon Lim, Dae-Won Kim, Jaesung Lee 2023	GTSNet (grouped temporal shift network)	UCI-HAR dataset WISDM dataset OPPORTUNITY dataset PAMAP2 dataset	95.7% , 88.6% , 87.4% , 76.2%
13	Action Recognition with Dynamic Image Networks	Hakan Bilen, Basura Fernando, Efstratios Gavves, and Andrea Vedaldi 2018	ResNeXt-50	UCF101 HMDB51	95% 74.5%
14	DCNN-Based Elderly Activity Recognition Using Binary Sensors	Munkhjargal Gochoo, Tan-Hsu Tan, Shih-Chia Huang, Shing-Hong Liu, Fady S. Alnajjar 2017	binary sensors (PIR sensor and door sensor) and Deep Convolutional Neural Network (DCNN)	Center for Advanced Studies in Adaptive Systems (CASAS)	98.78%
15	Human activity recognition using marine predators algorithm with deep learning	Ahmed M. Helmi , Mohammed A.A. Al qaness , Abdelghani Da hou, Mohamed Abd Elaziz 2023	residual convolutiona l network and a recurrent neural network (RCNN-BiGRU).and marine predator algorithm (MPA)	Opportunity, PAMAP2, and UniMiB-SHAR	
16	Ambient intelligence-based multimodal human action recognition for autonomous systems	Vidhi Jain, Gaurang Gupta, Megha Gupta, Deepak Kumar Sharma , Uttam Ghosh 2022	Hybrid Random Forest Bi-Convolutional Recurrent Neural Network(HRF Bi-CRNN)	SPHERE	94.2%
17	Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems	Muhammad Muaaz, Ali Chelli, Martin Wulf Gerdes, Matthias Patzold 2021	convolutional neural network (CNN)	Wi-Fi CSI dataset	97.78%
18	Multimodal Human Activity Recognition for Smart Healthcare Applications	Md. Milon Islam, Sheikh Nooruddin, and Fakhri Karray 2022	Convolutional Neural Networks (CNNs) with Convolutional Long Short Term Memory (ConvLSTM)	UP-Fall detection dataset	97.61%
19	Device-Free Human Activity Recognition Based on Dual-Channel Transformer Using WiFi Signals	Zhihao Gu, Taiwei He, Ziqi Wang, and Yuedong Xu 2022	Convolutional Neural Networks (CNN)	CSI dataset	93.3%
20	Smart Devices Based Multisensory Approach for Complex Human Activity Recognition	Muhammad Atif Hanif, Tallha Akram, Aamir Shahzad,Muhammad Attique Khan, Usman Tariq, Jung-In Choi, Yunyoung Nam, and Zanib Zulfiqar 2022	Naive Bayes (NB), K-Nearest Neighbors (KNN) and Neural Network (NN)	“Linear data collector V2	99.34%
21	A novel human activity recognition architecture: using residual inception ConvLSTM layer	Sarah Khater, Mayada Hadhoud and Magda B. Fayek 2022	residual inception convolutional recurrent layer, ResIncConvLSTM, and ConvLSTM lay	KTH and Weizmann datasets and UCF Sports Action dataset.	99.5% 99.7% 99.9%

22	SDHAR-HOME: A Sensor Dataset for Human Activity Recognition at Home	Raúl Gómez Ramos , Jaime Duque Domingo , Eduardo Zalama , Jaime Gómez-García-Bermejo and Joaquín López 2022	Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM) or Gated Recurrent Unit networks (GRU).	SDHAR-HOME	90.91% for user 1 and 88.29% for user 2 GRU
23	Acoustic- and Radio-Frequency-Based Human Activity Recognition	Masoud Mohtadifar , Michael Cheffena and Alireza Pourafzal 2022	hybrid acoustic- and radio-based method with Multi-Layer Perceptron (SVM) ,Random Forest, Extremely Randomized Trees (ERT), K-Nearest Neighbors (KNN), and Gradient Tree Boosting (GTB)	Real dataset	98%
24	Daily Human Activity Recognition Using Non-Intrusive Sensors	Raúl Gómez Ramos , Jaime Duque Domingo , Eduardo Zalama and Jaime Gómez-García-Bermejo 2021	bidirectional LSTM networks	CASAS	95.42%
25	Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals	Mingqi Lu, Yaocong Hu ,Xiaobo Lu 2019	deformable and dilated Faster R-CNN (DD-RCNN).	SEU-real-driving and Kaggle-driving dataset	92.1% for SEU-real-driving 86.0% for Kaggle-driving dataset
26	Real-Time Action Recognition System for Elderly People Using Stereo Depth Camera	Thi Thi Zin , Ye Htet , Yuya Akagi , Hiroki Tamura , Kazuhiro Kondo , Sanae Araki and Etsuo Chosa 2021	Stereo Depth Camera	dataset collected in the elder care center.	95.6%
27	A New Deep-Learning Method for Human Activity Recognition	Roberta Vrskova , Patrik Kamencay , Robert Hudec and Peter Sykora 2023	three-dimensional convolutional neural networks (3DCNNs) and Convolutional Long Short-Term Memory (ConvLSTM)	LoDVP Abnormal Activities , UCF50 and MOD20 dataset	89.12% for LoDVP, (UCF50mini) and MOD20 83.89% and 87.76%,
28	A deeply coupled ConvNet for human activity recognition using dynamic and RGB images	Tej Singh, Dinesh Kumar Vishwakarma 2020	CNN-Bi-LSTM bi-directional long short-term memory and ConvNet	The SBU Interaction, MIVIA Action, MSR Daily Activity 3D, MSR Action Pairs 3D dataset	98.3% For MSR Action Pairs 3D, 99.4% for MIVIA Action, 98.7% for SBU Interaction, 94.3% for MSR Daily Activity
29	Swimming Motion Analysis and Posture Recognition Based on Wearable Inertial Sensors	Zhelong Wang, Xin Shi, Jiaxin Wang,Fengshan Gao, Jie Li, Ming Guo, Hongyu Zhao and Sen Qiu 2019	wearable inertial sensors and HMM	Real dataset collection	94:38% for acceleration and gyroscope data, 91:84% for magnetometer and acceleration data and 86:83% for gyroscope and magnetometer data
30	Efficient activity recognition using lightweight CNN and DS-GRU network for surveillance applications	Amin Ullah, Khan Muhammad, Weiping Ding, Vasile Palade, Ijaz Ul Haq, Sung Wook Baik 2021	lightweight CNN model , MOSSE tracker and DS-GRU network	UCF-101, UCF-50, HMDB51, Hollywood2 and YouTube Actions	82.1% for UCF-101, 88.6% for UCF-50, 60.7% for HMDB51, 65.7% for Hollywood2 and 91.4% for YouTube Actions

31	Fusion Models for Generalized Classification of Multi-Axial Human Movement: Validation in Sport Performance	Rajesh Amerineni, Lalit Gupta , Nathan Steadman, Keshwyn Annauth , Charles Burr, Samuel Wilson, Payam Barnaghi and Ravi Vaidyanathan 2021	dynamic time warping (DTW) and convolutional neural networks (CNNs)	The eighteen-class boxing data	95%
32	ViCo-MoCo-DL: Video Coding and Motion Compensation Solutions for Human Activity Recognition Using Deep Learning	TAMER SHANABLEH 2023	Video Coding and Motion Compensation with Deep Learning	HMDB51, JHMDB and UCF11	78.5% for jHMDB dataset, 71.4% for HMDB51 dataset and 97% for UCF11 dataset
33	SlowFast Based Real-Time Human Motion Recognition with Action Localization	Gyu-Il Kim, Hyun Yoo and Kyungyong Chung 2023	YOLO and DeepSORT	AI Hub's abnormal action image data	70.74%
34	Human motion recognition based on Nano-CMOS Image sensor	Shangbin Li and Yu Liu 2023	Nano complementary metal oxide semiconductor (CMOS) image sensor	C-MHAD and MSR Daily Activity 3D data set	92%
35	Human motion recognition based on passive RFID and multi-model fusion	Xu Yang, Wenchao Luo, Xiaofeng An 2023	passive RFID and multi-model fusion.	The RSSI motion data	97.29%
36	Human Actions Recognition Based on 3D Deep Neural Network	Fadwa Al_Azzo, Chunbo Bao, Arwa Mohammed Taqi, Mariofanna Milanova, Nabeel Ghassan 2017	3D deep neural network (3D DNN)	KTH, Weizmann, and UCF101	98.90% for KTH 97.02 %.for Weizmann and 100%for UCF 101
37	3D Human Action Recognition using Hu Moment Invariants and Euclidean Distance Classifier	Fadwa Al-Azzo,Arwa Mohammed Taqi and Mariofanna Milanova 2017	Hu moment invariants HMI algorithm	KTH, and UCF101	93.4% for KTH 92.11%for UCF101

References

[1] S. Angerbauer, A. Palmanshofer, S. Selinger, and M. Kurz, "applied sciences Comparing Human Activity Recognition Models Based on Complexity and Resource Usage," 2021.

[2] and O. U. P. Turaga, R. Chellappa, V. S. Subrahmanian, "Machine recognition of human activities: A survey, Circuits and Systems for Video Technology," *IEEE Trans.*, vol. 18, no. 11, pp. 1473–1488, 2008.

[3] Z. Hussain, M. Sheng, and W. E. Zhang, "Different Approaches for Human Activity Recognition: A Survey," pp. 1–28, 2019, doi: 10.1016/j.jnca.2020.102738.

[4] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 42, no. 6. pp. 790–808, 2012, doi: 10.1109/TSMCC.2012.2198883.

[5] D. Bouchabou, S. M. Nguyen, C. Lohr, B. Leduc, and I. Kanellos, "A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning," *Sensors*, vol. 21, no. 18. MDPI, Sep. 01, 2021, doi:

10.3390/s21186037.

[6] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, “Sensor-based and vision-based human activity recognition: A comprehensive survey,” *Pattern Recognit.*, vol. 108, Dec. 2020, doi: 10.1016/j.patcog.2020.107561.

[7] C. Kim and W. Lee, “Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data,” *Appl. Sci.*, vol. 13, no. 8, 2023, doi: 10.3390/app13084961.

[8] M. M. and A. A. B. and M. Y. and R. B. Gopaluni, “A Vision-based Deep Learning Platform for Human Motor Activity Recognition,” pp. 1–4, 2023, doi: 10.1109/MOCAST57943.2023.10176420.

[9] L. Xia, C. Chen, and J. K. Aggarwal, “View Invariant Human Action Recognition Using Histograms of 3D Joints The University of Texas at Austin.”

[10] M. G. Morshed, T. Sultana, A. Alam, and Y. K. Lee, “Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities,” *Sensors*, vol. 23, no. 4. MDPI, Feb. 01, 2023, doi: 10.3390/s23042182.

[11] C. Schüldt, I. Laptev, and B. Caputo, “Recognizing human actions: A local SVM approach,” in *Proceedings - International Conference on Pattern Recognition*, 2004, vol. 3, pp. 32–36, doi: 10.1109/ICPR.2004.1334462.

[12] J. Liu, A. Shahroudy, M. L. Perez, G. Wang, L. Duan, and A. K. Chichung, “NTU RGB + D 120 : A Large-Scale Benchmark for 3D Human Activity Understanding,” no. 1, p. 120.

[13] Institute of Electrical and Electronics Engineers, *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on : date, 13-18 June 2010.* .

[14] W. Li, Y. Wong, A.-A. Liu, Y. Li, Y.-T. Su, and M. Kankanhalli, “Multi-Camera Action Dataset for Cross-Camera Action Recognition Benchmarking,” Jul. 2016, doi: 10.1109/WACV.2017.28.

[15] IEEE Staff and IEEE Staff, *2012 IEEE Conference on Computer Vision and Pattern Recognition.* .

[16] S. Singh, S. A. Velastin, and H. Ragheb, “MuHAVi: A multicamera human action video dataset

for the evaluation of action recognition methods," in *Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 2010, pp. 48–55, doi: 10.1109/AVSS.2010.63.

[17] K. K. Reddy and M. Shah, "Recognizing 50 human action categories of web videos," *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 971–981, 2013, doi: 10.1007/s00138-012-0450-4.

[18] IEEE Computer Society., *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) : date, 7-12 June 2015.* .

[19] W. Kay *et al.*, "The Kinetics Human Action Video Dataset," May 2017, [Online]. Available: http://arxiv.org/abs/1705.06950.

[20] Institute of Electrical and Electronics Engineers, IEEE International Conference on Computer Vision 13 2011.11.06-13 Barcelona, and ICCV 13 2011.11.06-13 Barcelona, *IEEE International Conference on Computer Vision (ICCV), 2011 6 - 13 Nov. 2011, Barcelona, Spain*. .

[21] *Computer Vision and Pattern Recognition, 2009, CVPR 2009, IEEE Conference on : dates: 20-25 June 2009.* IEEE, 2009.

[22] G. Morshed, T. Sultana, A. Alam, and Y. Lee, "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities," pp. 1–40, 2023.

[23] M. G. Amin and B. Erol, "Understanding Deep Neural Networks Performance for Radar-based Human Motion Recognition," *2018 IEEE Radar Conf.*, pp. 1461–1465, 2018, doi: 10.1109/RADAR.2018.8378780.

[24] C. C. Vu, "Human Motion Recognition by Textile Sensors Based," 2018, doi: 10.3390/s18093109.

[25] F. Hafizhelmi, K. Zaman, H. Ali, A. A. Shafie, and Z. I. Rizman, "Efficient Human Motion Detection with Adaptive Background for Vision- Efficient Human Motion Detection with Adaptive Background for Vision-Based Security System," no. June, 2017, doi: 10.18517/ijaseit.7.3.1329.

[26] Y. Shao, S. Guo, L. Sun, and W. Chen, "Human Motion Classification Based on Range Information with Deep Convolutional Neural Network," pp. 1520–1524, 2017, doi:

10.1109/ICISCE.2017.317.

[27] S. U. Park, J. H. Park, M. A. Al-masni, M. A. Al-antari, Z. Uddin, and T. Kim, “A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services,” *Procedia - Procedia Comput. Sci.*, vol. 100, pp. 78–84, 2016, doi: 10.1016/j.procs.2016.09.126.

[28] M. Gong and Y. Shu, “Real-time Detection and Motion Recognition of Human Moving Objects Based on Deep Learning and Multi-scale Feature Fusion in Video,” 2020, doi: 10.1109/ACCESS.2020.2971283.

[29] F. Zhang, T. Y. Wu, J. S. Pan, G. Ding, and Z. Li, “Human motion recognition based on SVM in VR art media interaction environment,” 2019.

[30] D. Hendry, K. Chai, A. Campbell, L. Hopper, P. O’Sullivan, and L. Straker, “Development of a Human Activity Recognition System for Ballet Tasks,” *Sport. Med. - Open*, vol. 6, no. 1, 2020, doi: 10.1186/s40798-020-0237-5.

[31] V. C. Mariani and S. Coelho, “Video-Based Human Activity Recognition Using Deep Learning Approaches,” pp. 1–15, 2023.

[32] V. Chavan, “ScienceDirect ScienceDirect Real-Time Deep Learning Approach for Pedestrian Detection and Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Suspicious Activity Activity Recognition Recognition,” *Procedia Comput. Sci.*, vol. 218, pp. 2438–2447, 2023, doi: 10.1016/j.procs.2023.01.219.

[33] K. Guo, P. Wang, P. Shi, and C. He, “applied sciences A New Partitioned Spatial – Temporal Graph Attention Convolution Network for Human Motion Recognition,” 2023.

[34] J. Park, W. Lim, D. Kim, and J. Lee, “Engineering Applications of Artificial Intelligence GTSNet : Flexible architecture under budget constraint for real-time human activity recognition from wearable sensor,” *Eng. Appl. Artif. Intell.*, vol. 124, no. May 2022, p. 106543, 2023, doi: 10.1016/j.engappai.2023.106543.

[35] H. Bilen, B. Fernando, E. Gavves, and A. Vedaldi, “Action Recognition with Dynamic Image Networks,” no. December, 2018, doi: 10.1109/TPAMI.2017.2769085.

[36] T. Tan, “DCNN-Based Elderly Activity Recognition Using Binary Sensors,” no. February 2018, 2017, doi: 10.1109/ICECTA.2017.8252040.

[37] M. A. A. A. Ahmed M. Helmi, “Human activity recognition using marine predators algorithm with deep learning,” *ScienceDirect*, 2023, [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167739X23000134#preview-section-snippets.

[38] V. Jain, G. Gupta, M. Gupta, D. Kumar, and U. Ghosh, “Ambient intelligence-based multimodal human action recognition for autonomous systems,” *ISA Trans.*, vol. 132, pp. 94–108, 2023, doi: 10.1016/j.isatra.2022.10.034.

[39] M. Muaaz, A. Chelli, M. Wulf, and G. Matthias, “Wi-Sense : a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems,” pp. 163–175, 2022.

[40] M. Islam, S. Nooruddin, and F. Karray, “Multimodal Human Activity Recognition for Smart Healthcare Applications,” *2022 IEEE Int. Conf. Syst. Man, Cybern.*, no. November, pp. 196–203, 2022, doi: 10.1109/SMC53654.2022.9945513.

[41] Z. Gu, T. He, Z. Wang, and Y. Xu, “Device-Free Human Activity Recognition Based on Dual-Channel Transformer Using WiFi Signals,” vol. 2022, 2022.

[42] M. A. Hanif *et al.*, “Smart Devices Based Multisensory Approach for Complex Human Activity Recognition,” 2022, doi: 10.32604/cmc.2022.019815.

[43] S. Khater, M. Hadhoud, and M. B. Fayek, “Open Access A novel human activity recognition architecture : using residual inception ConvLSTM layer,” vol. 0, 2022.

[44] R. G. Ramos, J. D. Domingo, E. Zalama, J. Gómez-garcía-bermejo, and J. López, “SDHAR-HOME : A Sensor Dataset for Human Activity Recognition at Home,” pp. 1–27, 2022.

[45] M. Mohtadifar, M. Cheffena, and A. Pourafzal, “Acoustic- and Radio-Frequency-Based Human Activity Recognition,” 2022.

[46] D. Human, A. Recognition, and N. Sensors, “Non-Intrusive Sensors,” pp. 1–19, 2021.

[47] M. Lu, Y. Hu, and X. Lu, “Driver action recognition using deformable and dilated faster R-

CNN with optimized region proposals," 2019.

[48] S. D. Camera, "Real-Time Action Recognition System for Elderly People Using Stereo Depth Camera," 2021.

[49] R. Vrskova, P. Kamencay, R. Hudec, and P. Sykora, "A New Deep-Learning Method for Human Activity Recognition," 2023.

[50] T. Singh and D. Kumar, "A deeply coupled ConvNet for human activity recognition using dynamic and RGB images," *Neural Comput. Appl.*, vol. 0123456789, 2020, doi: 10.1007/s00521-020-05018-y.

[51] Z. Wang *et al.*, "Swimming Motion Analysis and Posture Recognition Based on Wearable Inertial Sensors," pp. 3371–3376, 2019.

[52] A. Ullah, K. Muhammad, W. Ding, V. Palade, and I. Ul, "Efficient activity recognition using lightweight CNN and DS-GRU network for surveillance applications," vol. 103, 2021, doi: 10.1016/j.asoc.2021.107102.

[53] R. Amerineni *et al.*, "Fusion Models for Generalized Classification of Multi-Axial Human Movement : Validation in Sport Performance," 2021.

[54] T. Shanableh and S. Member, "ViCo-MoCo-DL : Video Coding and Motion Compensation Solutions for Human Activity Recognition Using Deep Learning," *IEEE Access*, vol. 11, no. July, pp. 73971–73981, 2023, doi: 10.1109/ACCESS.2023.3296252.

[55] G. Kim, H. Yoo, and K. Chung, "SlowFast Based Real-Time Human Motion Recognition with Action Localization," 2023, doi: 10.32604/csse.2023.041030.

[56] S. Li and Y. Liu, "Human motion recognition based on Nano-CMOS Image sensor," vol. 20, no. January, pp. 10135–10152, 2023, doi: 10.3934/mbe.2023444.

[57] X. Yang, W. Luo, and X. An, "passive RFID and multi-model fusion," no. July, 2023, doi: 10.1117/12.2685530.

[58] L. Rock, L. Rock, and L. Rock, "Human Actions Recognition Based on 3D Deep Neural Network," no. March, 2017, doi: 10.1109/NTICT.2017.7976123.

[59] F. Al-azzo and A. M. Taqi, “3D Human Action Recognition using Hu Moment Invariants and Euclidean Distance Classifier,” vol. 8, no. 4, pp. 13–21, 2017.

التطورات في التعرف على النشاط البشري: مسح شامل للمناهج القائمة على الاستشعار والرؤية

الخلاصة: *زادت أهمية التعرف على النشاط البشري (HAR) في السنوات الأخيرة بسبب تطبيقاته واسعة النطاق في مجالات مثل الرعاية الصحية والأمن والمراقبة والترفيه والإعدادات الذكية. أحد التحديات الكبيرة في رؤية الكمبيوتر هو التعرف الآلي والدقيق على الأفعال البشرية. يعرض هذا المسح آخر الأعمال ذات الصلة التي تم إجراؤها خلال الأعوام 2015-2023 في مجالات مختلفة للتعرف على النشاط البشري. ويقدم تصنيف المنهجيات الأساسية HAR. بشكل عام، يتم تصنيف مناهج HAR إلى مجموعتين أساسيتين: HAR القائمة على المستشعر والرؤية. يعتمد هذا التصنيف على نوع البيانات التي تم إنشاؤها وبيئة النظام نفسها. بناءً على دراستنا، تعد الهواتف الذكية، والساعات الذكية، ومقاييس التسارع، والجيروسكوبات، وأساور الذراع، كلها أمثلة على التقنيات المعتمدة على أجهزة الاستشعار. أما بالنسبة للتقنيات المعتمدة على الرؤية، فهناك الكاميرات، ومايكروسوفت كينيكت، والكاميرات الحرارية. يتم أيضًا تصنيف مجموعات البيانات العامة المشار إليها في الأوراق الأكاديمية بناءً على نوع النشاط: الإجراءات الفردية والسلوكيات والتفاعلات والأنشطة الجماعية. وبعد ذلك، يتم عرض إجراءات المعالجة المسبقة وهندسة الميزات. أخيرًا، هذه المراجعة قادرة على تقديم بعض مفاهيم الدراسة التي تبحث وتحلل HAR*

الكلمات المفتاحية: التعرف على النشاط البشري (HAR)، الذكاء الاصطناعي (AI)، مجموعة البيانات الشائعة المستندة إلى الرؤية والمستشعرة