Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-15443-5

A review of vision-based indoor HAR: state-of-the-art, challenges, and future prospects

Geetanjali Bhola[1] · **Dinesh Kumar Vishwakarma**[1]

Received: 28 December 2021 / Revised: 10 February 2023 / Accepted: 18 April 2023

Abstract
With the advent of technology, we are getting more comfortable with the use of gadgets, cameras, etc., and find Artificial Intelligence as an integral part of most of the tasks we perform throughout the day. In such a scenario, the use of cameras and vision-based sensors comes as an escape from many real-time problems and challenges. One major application of these vision-based systems is Indoor Human Activity Recognition (HAR) which serves in a variety of scenarios ranging from smart homes, elderly care, assisted living, and human behavior pattern analysis for identifying any abnormal behavior to abnormal activity recognition like falling, slipping, domestic violence, etc. The effect of HAR in real time has made the area of indoor activity recognition a more explored zone by the industrial segment to attract users with their products in multiple domains. Hence, considering these aspects of HAR, this work proposes a detailed survey on indoor HAR. Through this work, we have highlighted the recent methodologies and their performance in the field of indoor activity recognition. We have also discussed- the challenges, detailed study of approaches with real-world applications of indoor-HAR, datasets available for indoor activity, and their technical details in this work. We have proposed a taxonomy for indoor HAR and highlighted the state-of-the-art and future prospects by mentioning the research gaps and the shortcomings of recent surveys with respect to our work.

Keywords Human activity recognition · Assistive technology · Assisted living · Elderly care · Indoor activity · Patient care · Real-time HAR

1 Introduction

Human activity recognition has been studied, reviewed, and discussed in some of the previous works [13, 15, 24, 47, 85, 105, 132, 160, 175, 190, 216]. Human activity recognition has been under constant limelight in the area of video analysis technology due to the rising needs from many key areas like entertainment environments,

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Bawana Road, Delhi 11042, India

human-computer interaction (HCI), surveillance environments, and healthcare systems. Its application in the surveillance environment involves the detection of abnormal activities or dangerous behaviors which can alert the related authorities. Similarly, in an entertainment environment, activity recognition can be used to improve human-computer interaction (HCI), for e.g., by embedding AR (augmented reality) features in real-time applications [27]. Furthermore, activity recognition can help in the automatic recognition of patient's actions to facilitate the rehabilitation processes in a healthcare system. Applications in the area of human interaction [61], pedestrian traffic [93], home abnormal activity [63], human gestures [33], ballet activity [29], tennis activity [108, 203], sports activity [133, 163] simple actions [1, 25, 60, 91, 113, 125, 145, 166–168, 187] and healthcare applications [2, 20, 64, 71–74, 77, 78, 81, 114, 118, 120, 124, 129, 131] are a few examples of HAR.

Indoor-HAR, which can be considered as one of the emerging areas of HAR, deals with the recognition of activities limited to an indoor environment like homes, gyms, sports clubs, hospitals, schools, corridors, parking lots, etc. The area of Indoor-HAR is different in terms of the challenges and applications it has. For example, we can relate activities like walking, moping, and wandering to a school's corridor. Whereas jumping, weight lifting, running, etc. can be a part of gym activities. Activities inside a house, hospital, sports complex, gym, or any closed environment can be limited compared to outdoor activities. These activities can be basic ones like walking, sitting, sleeping, studying, cooking, moping, drinking water, etc. or complex ones like "taking medication" which can involve a series of simple activities like "opening the pill box", and "drinking water". Similarly, another complex activity can be "praying" which can further involve activities like "sitting" or "standing", "holding a book(studying)", "ringing a bell" or "lighting a candle" or "lamp".

Certain activities may be simple but the pose of the person performing the activity and the orientation may make them complex for the machine to recognize. An activity like "drinking water" while sitting, standing or walking may generate different recognition rates. Also, a few activities may have certain similarities while performing due to which the results can be incorrect. For e.g., jumping and dancing/doing aerobics may have similar movements which may be ambiguous for the system.

In order to recognize activities limited to some indoor environments, we cannot use conventional deep learning methods as there are different challenges involved, based on the environment. As in the case of indoor HAR, the activity set differs from one environment to the other, one model cannot fit in every scenario.

In other words, the use of one type of HAR model meant for general may not be suitable for a specific indoor environment. So, indoor HAR is a different track that requires the problem of activity recognition to be considered differently as per its application area involved. Through this survey, we have tried to introduce Indoor-HAR as a new and emerging area of research. Through indoor HAR, concepts like smart homes, ADL(Activities of Daily Living), automated patient care, and elderly care can be strengthened and practically have a lot of research potential to be looked into. In this survey, we have discussed recent developments in the field of indoor HAR, different approaches, datasets, the use of indoor HAR for human behavior analysis, and the real-time application of indoor HAR.

The main contributions of this work are as follows:

- It outlines the comparison of previous surveys in indoor HAR, which includes its merits and shortcomings.
- Deliberates the latest and important developments recently reported covering both the general aspects of HAR and the specific vision-based indoor-HAR systems.

- Elaborates the scope and wide application area of indoor HAR, discussing various existing and suggested research works which can help in real-life situations
- Proposes a taxonomy for indoor-HAR on the basis of input methodology, datasets, and challenges which one by one explores the role of each of these parameters and how are they connected to HAR and indoor-HAR.
- Highlights the HAR and specifically indoor HAR datasets which includes the technical details like dimensionality, fps, number of activities etc.
- The methodology included for indoor HAR is highlighted which includes the use of handcrafted features and automatic feature learning in HAR
- Summarises the detailed findings of various state-of-art, challenges and hurdles involved at different levels of processing and at different stages of research for any indoor-HAR system.

The steps involved in HAR are shown in Fig. 1. Depending on the nature of the data, it can be vision-based or sensor-based.

Previously, there have been some surveys [6, 24, 34, 84, 141] in the field of HAR which have been summarised below in Table 1.

The ultimate aim of this work is to bridge the gap between existing parameters and the real-time applications of indoor HAR. With this survey, we wish to enhance the usability of existing works for more practical and real-time scenarios.

The area of indoor-HAR is vast and can have a variety of scope for future work. Some of the application areas are discussed here.

The sections are explained as follows: section 2 gives the literature review, and section 3 discusses the input methodology for the various research techniques, which are RFID/sensor-based and vision based. Section 4 discusses datasets; section 5 discusses the approaches and human body representation. Section 6 discusses the challenges and section 7 gives conclusions and future scope.

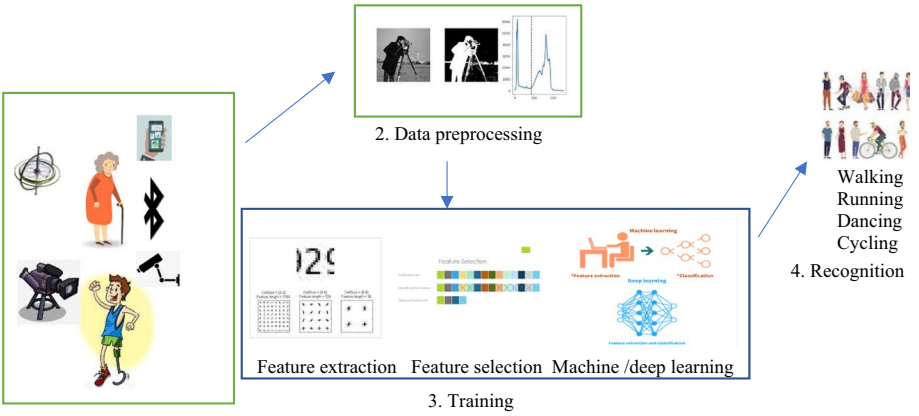

Fig. 1 Steps for HAR i) data collection ii) data pre-processing iii) training iv) recognition

Table 1 Recent surveys in the field of HAR

Ref.	Year	Merits	Shortcomings
[34]	2016	Comprehensive state-of-art survey for different phases of HAR	No detail about real-time feasibility is covered.
[141]	2020	Covers sensor-based and vision-based HAR; Discusses datasets, methods involved in data processing and feature engineering, future direction	Sensor and vision-based methods are discussed in one paper and an overall conclusion on the suitability for real-time application is not discussed.
[24]	2020	Provides an up-to-date analysis of vision-based HAR-related literature and recent progress in the field. Highlights the challenges and future prospects	Gives overall broader picture of different aspects related to HAR and not specific to one domain
[84]	2021	An in-depth and comprehensive survey on HAR with recently developed deep learning methods	Requirement of lot of labelled data; non-existence of universal activity recognition model with multimodal signals and data.
[6]	2021	Surveys two distinct domains of IoT and DL and their amalgamation for indoor HAR and position/ tracking system	IoT has challenges like security, connectivity and compatibility which have not been covered [181]
[177]	2022	Reviews the different modalities in HAR	Only modalities have been discussed. Other components are not reviewed in detail.
Our work	–	Talks specifically of indoor-HAR with a vision-based scenario in particular. Proposed indoor-HAR taxonomy based on different parameters associated Lists datasets explicitly for indoor HAR	Role of sensor-based indoor HAR is not covered in detail

2 Related work

The area of HAR has been popular for many years now and a lot has been achieved in this area. This section will discuss various applications of Indoor-HAR and their related work and the recent research that has been done in that area involving both vision-based and sensor-based input.

2.1 Elderly care system

There is a need for extra care for the elderly and people with special needs as their number is increasing and there is a shortage of care staff. During recent times of COVID-19, we came across many instances where family members were not there to look after the elderly and they died in isolation. Some countries also faced a dire shortage of nurses and caretakers. This is a consistent problem where the elderly lives alone. The monitoring is difficult. Wearable sensors are not very efficient in the case of old people as they tend to forget to wear them. Sometimes, the uneasiness caused while wearing them can cause hurdles in totally depending on such sensors for their activity monitoring. Also, the statistics cannot be as impactful as watching a person doing some activity. In such cases, a vision-based HAR can provide a solution for the challenges and problems related to wearable sensors. Vision-based methods deploy a camera for recognising the activities of a person. Different research works based in the same field are [69, 106]. Recent work includes [153] which proposes unsupervised HAR with skeletal Graph Laplacian Invariance.

The daily activities can be detected and a pattern can be observed which can help in understanding their needs better. Apart from regular monitoring, systems for special alerts can be designed which can help in situations like falls, slipping, painful walking or limping and some abnormal activity with respect to the environment. The risk of falling is one of the most prevalent problems faced by elderly individuals. A study published by the World Health Organization [198] estimates that between 28% and 35% of people over 65 years old suffer at least one fall each year, and this figure increases to 42% for people over 70 years old. According to the World Health Organization, falls represent greater than 50% of elderly hospitalizations and approximately 40% of the non-natural mortalities for this segment of the population. A model which is trained for the detection of "fall" [72, 176, 206, 211] might be used for alerting the neighbors or other members of the house about the same. In the survey [212], vision-based fall detection systems were discussed. Vision-based approaches of fall detection were divided into four categories, namely individual single RGB cameras, infrared cameras, depth cameras, and 3D-based methods using camera arrays. Other works based on fall detections are [67, 87, 149, 207]. Figure 2 gives an idea of how a vision-based system can be used for monitoring the elderly [35].

2.2 Patient care system

Vision sensors(cameras) can be deployed to monitor a patient who may need constant observation or is going under treatment or needs post-recovery care. Apart from vital signs which are monitored for such patients, a vision-based system may be helpful in case or abnormalities or medical emergencies. Capturing the visual cues, facial gestures or

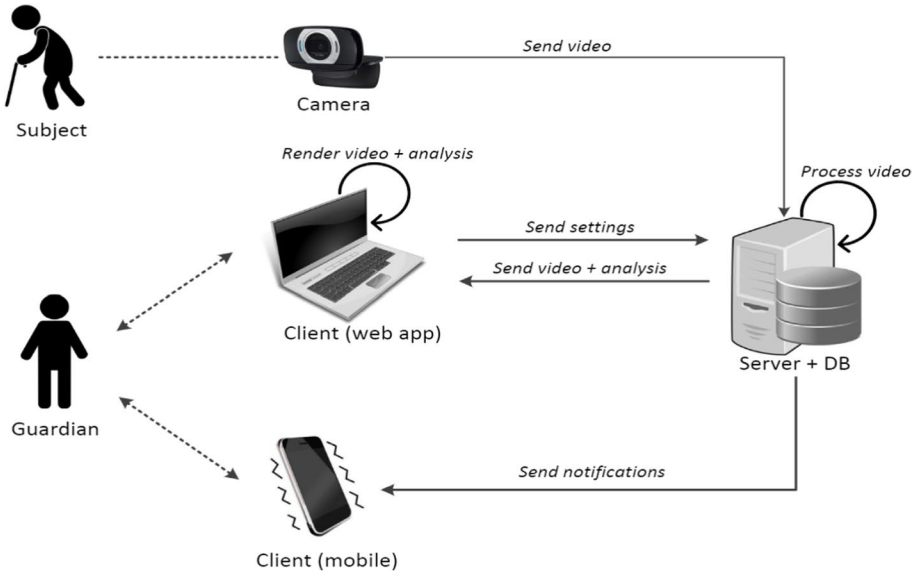

Fig. 2 Elderly monitoring vision-based system [35]

detection of motion or activity of such patients can help in alerting the concerned team which can be significant in the recovery of a patient.

2.3 Physical fitness

Intelligent systems have been built over the years which have been designed to benefit us in our fitness domain. Multiple wearable fitness trackers have boomed the market over the years. These trackers have sensors that record our data and instruct us about better workouts, balanced eating, and sensible living. Research on pose estimation directs naturally to human activity recognition [48, 157]. Smartphones have also given a new shape to the fitness model. The ubiquity of smartphones together with their ever-growing computing, networking, and sensing powers have been changing the landscape of people's daily life [76, 191].

Wearable sensors and smartphones have opened a pool of options in the fitness domain for research but these methods may suffer from problems like interference, the privacy of data, unease of wearing the sensors, and skin irritation [197]. Vision-based systems can be effective as they offer an unobtrusive solution for the monitoring and diagnosis of the person. Google's on-device, real-time body pose tracking with MediaPipe BlazePose is a great research example that is making use of human body pose and can be used to build fitness and yoga trackers [82]. Figure 3 shows the screenshots of the real-time results captured from the same.

2.4 Abnormal human activity recognition

Abnormal activity means any irregularity in the set of activities belonging to an environment. When we talk of abnormal activities in an indoor environment, we mean any activity

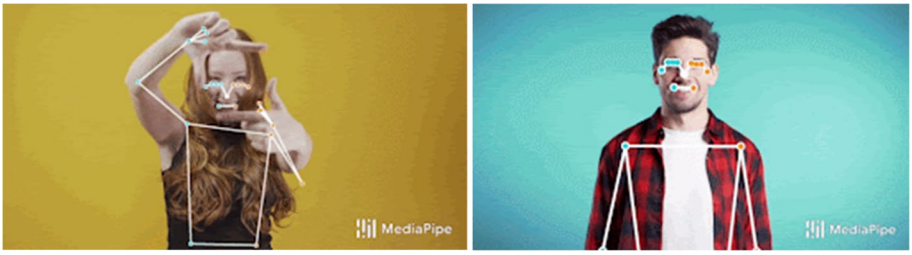

Fig. 3 Google's real-time pose estimation in fitness domain [82]

which can be a matter of concern or may require immediate attention [17]. We come across cases of domestic violence, child sexual and physical abuse, and even physical abuse of the elderly at home. Researchers have also worked on providing a solution to this problem [185]. As per the reports [49], worldwide, approximately 20% of women and 5–10% of men are sexually abused in their childhood. In India, 2 out of every 3 children are physically abused and every second child is a victim of emotional abuse. In general, these malpractices are never done in the open, and thus saving such children is an essential step toward building a healthy and happy future. Indoor activity recognition may be vision-based or wi-fi signal based. Systems are built for accurately detecting abnormal activities on commercial-off-the-shelf (COTS) IEEE 802.11 devices [55, 215]. This system makes use of IoT and wi-fi signals for any abnormal indoor activity recognition.

2.5 Human abnormal behavior analysis

Our society suffers from a lack of awareness about mental health. It is taboo to talk about mental health issues. Another useful application for indoor HAR can be for analyzing human behavior by studying the pattern of indoor activities. A sick person may have a different activity pattern than a healthy person. He may be less active and may not be performing his activities like on usual days. This can also help in detecting any mental illness like depression where a person may repeat an activity or may skip his daily activities. Research like this may be helpful for situations like the onset of depression.

Recent works involve the use of smartphone data collection of the sensory sequence which is collected from people performing their daily tasks. Later, the cycle detection algorithm helps in segmenting the data sequence for activity [41].

Indoor-HAR taxonomy In the present era, human activity recognition [12, 34, 56, 65, 188, 189] in videos has become a prominent research area in the field of computer vision. Videos are used in daily living applications such as patient monitoring, object tracking, threat detection, security, and surveillance. In general, all HAR algorithms are categorized on various aspects from how the data acquisition is performed to what activities are considered. Figure 4 shows different facets of indoor HAR.

The taxonomy is based on parameters like the approach, data type, data sets, 2D or 3D representation of the input, etc. This survey also discusses the challenges and problems at different levels of HAR recognition for indoor activities. The survey covers the parameters shown in the figure and throws light on them one by one w.r.t the existing work in that field and the characteristics of the same.

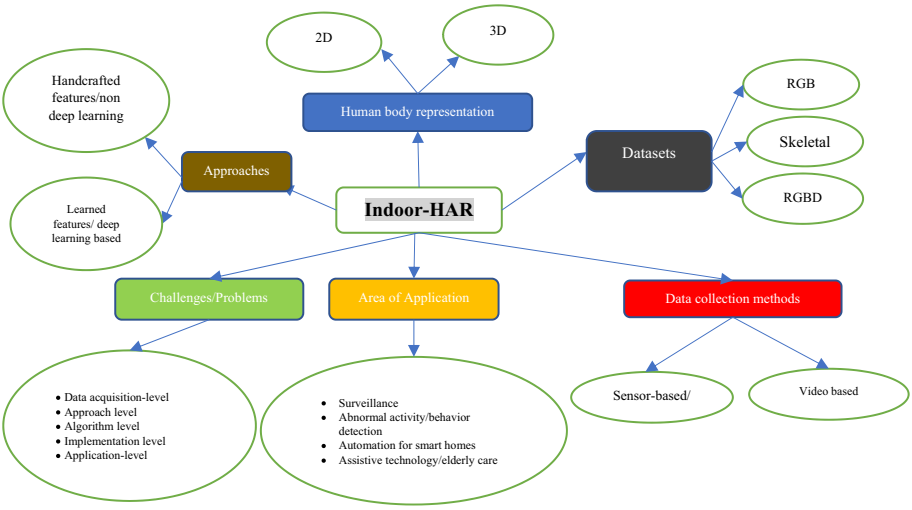

Fig. 4 Taxonomy of Indoor-HAR

3 Input methods: vision-based and sensor-based

Specifically, indoor HAR methods are classified into two main groups, which are sensor-based/RFID-based HAR and vision-based HAR, based on the generated data type. [97, 202] talks about another category called multimodal where sensor data, as well as visual data, are used to detect human activities.

Sensor technology has grown in multiple perspectives, including computational power, size, accuracy, and manufacturing costs. This has widened the band of portable devices which are making use of this technology to record data for activity recognition. Examples are wristbands, smartphones, etc. This growth and development of sensor technology have also fostered techniques for pervasive human tracking, silhouette tracking, detection of uncertain events [90, 98, 208], human motion observation, and emotion recognition in real environments [3, 99, 192].

Human tracking and activity recognition problems require feature extraction and pattern recognition techniques based on specific input data from innovative sensors (i.e., motion sensors and video cameras) [4, 5, 100, 152, 204]. Motion sensors-based activity recognition is based on classifying sensory data using one or more sensor devices. [39], proposed a complete review about the state-of-the-art activity classification methods using data from one or more accelerometers. In [102], the classification approaches are based on RFs features which classify five daily routine activities from Bluetooth accelerometer placed at breast of the human body, using a 319-dimensional feature vector. In [22], fast FFT and decision tree classifier algorithms are proposed to detect physical activity using biaxial accelerometers attached to different parts of the human body. However, these motion sensors-based approaches are not feasible methods for recognition due to the discomfort of the users to wear electronic sensors in their daily life. Also, combining multiple sensors for improvement in recognition performance causes a high computation load.

Recently, there has been a tremendous increase in the number of applications using closed-circuit television (CCTV) for monitoring and security purposes due to the evolution in CCTV technology which has resulted in better video quality, more straightforward

setup, lower cost, and secure communication. Although each type of sensor aims at specific services and applications, sensors generally collect raw data from their target ubiquitously, and general knowledge is acquired by analyzing the collected data. Human activity recognition or HAR, allows machines to analyze and comprehend several human activities from input data sources, such as sensors, and multimedia content. HAR is applied in surveillance systems which can be for home [21, 147], for mass crowd-monitoring [38, 95] and for detecting humans under distress using UAVs [134], behavior analysis, gesture recognition, patient monitoring systems, ambient assisted living (AAL), and a variety of healthcare systems that involve direct interaction or indirect interaction between human and smart devices. HAR for elderly care has been a key area of work [37, 101, 109]. Various opportunities can be provided by these sensor technologies which can improve the robustness of the data through which human activities can be detected and also provide services based on sensed information from real-time environments, such as cyber-physical-social systems [205]; there is also a type of magnetic sensors when embedded in a smartphone can track the positioning [19]. The above points can be summarised in Fig. 5. Table 2 lists recent works in sensor-based input and vision-based input with the technique used.

3.1 Vision-based input

With the ever-developing field of HAR, there is also a need of a more practical way of fetching the data. This is where vision-based HAR proves efficient as compared to sensor-based HAR. The sensor-based HAR involves the placement of sensors as per the activity which is difficult to plan whereas vision-based HAR is a compatible approach for real-time videos which are the most common source of input and data considering the increase of CCTVs and cameras. The vision-based HAR is an important research area of computer vision. Vision-based HAR helps us identify the activity being performed in the video/image. The task is cumbersome due to the challenges involved like cluttered background, shape variation of the objects/people involved, illumination differences, placement of camera(s), viewpoint, stationery or moving camera etc. These challenges are further dependent on the category of activity being considered. There are four major categories of activities: gestures, action, interaction, and group activity represented in Fig. 6. Indoor HAR can include activities from these categories.

[93] discusses components of an automated, "smart video" system to track pedestrians and detect situations where people may be in peril, as well as suspicious motion or

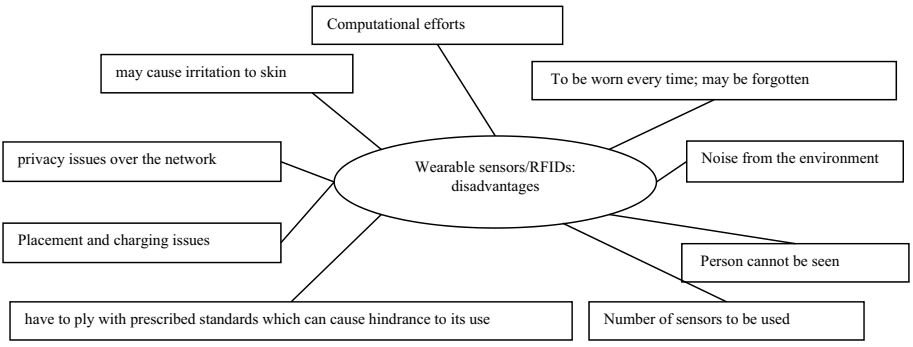

Fig. 5 Disadvantages of wearable sensors/RFIDs

Table 2 Recent work with sensor-based and vision-based inputs

Ref.	Year	Key contributions	Input type	Advantages	Disadvantages	Dataset used	Technique
[32]	2021	comprehensive dataset intended to evaluate passive Human Activity Recognition (HAR) and localization techniques with measurements obtained from synchronized Radio-Frequency (RF) devices and vision-based sensors	Sensor, Vision	Captures transitioning of activities too. Like walking to falling	Spectrograms may differ in most realistic scenarios as the human motions might not be restricted to a single aspect angle with respect to the radar. It could be due to the shadowing of some part of the human body if captured at a different angle	Gathered during experiment	ML techniques
[28]	2021	integrated system prototype that provides an efficient technological tool to caregivers operating promptly and ensures efficient performance throughout the entire healthcare system process	Sensor	Takes care of situations where the elderly are to be monitored every time without the use of expensive cameras etc	Sensor placements	Wireless Sensor Data Mining (WISDN) dataset in Jennifer R. Kwapisz et al. t. [153].	Localization using CNN
[214]	2021	Joint tracking and activity recognition in indoor environment using radar sensors	Radar sensors	Placement of sensors even in sensitive areas is not a privacy concern	Boundary error Data representation Imbalanced dataset Loss function	Collect own dataset using the setup uses K-fold method as dataset is limited	Deep learning
[146]	2021	(HAR) based on Radio Frequency energy harvesting (RFEH) as the harvested voltage signals of different human activities exhibit distinctive patterns.	Radio frequency	Accuracy and computational efficiency	Interference from other wifi devices Complex activities and variety of ways in which one may perform them.	Gathered dataset performance of the system by applying the four light-weight classifiers and calculating their accuracies on the collected datasets	Machine learning classifiers

Table 2 (continued)

Ref.	Year	Key contributions	Input type	Advantages	Disadvantages	Dataset used	Technique
[51]	2021	HAR for indoor using channel information of wifi signals	Wifi signals	Cost-effective as wifi devices are mostly already installed.	Interference in signal	Collected own data set	CNN
[182]	2021	Smart-home, smart-gym fitness tracker solution using a single mm-wave radar point cloud data	Radar sensors	The privacy of the user is intact. contactless, accurate, real-time fitness tracker system for indoor fitness activities, capable of running on edge devices for applications in IoT-connected healthcare	Interference/noise has not been commented upon	Data set collected from radar sensors	Deep learning
[170]	2021	Combination of acceleration audio and wifi round trip time localization to recognize indoor activities.	Wifi round trip audio acceleration	Audio related to activity gives another useful feature in recognition. Accuracy is consistent upon 12 activities	Limited activities, Maybe not suitable if activities with similar audio are tested, technology used is relatively new and hardware is not ubiquitous	Dataset created and model trained again and again for new activities	ML
[7]	2021	Gives a mathematical representation for human body as clusters of different points	Radio frequency	Time variant features and properties explored while representing human as cluster of scattered points	Signa based, stationary objects	CSI and IMU data were used as collected	ML

Table 2 (continued)

Ref.	Year	Key contributions	Input type	Advantages	Disadvantages	Dataset used	Technique
[83]	2015	online activity recognition system, which explores WiFi ambient signals for received signal strength indicator (RSSI) fingerprint of different activities.	Wifi signal data	Simple use. Fine average accuracy	context between persons and motions, context between locations and motions, and even context between emotions and motions can be explored as the current scenario is simple	Data collected using set up/ wifi	ML
[186]	2018	Uses video camera and radar sensors fusion for predicting activities	Radar sensors, video camera	Fusion of radar and video with deep learning gives good accuracy	Computation can be worked upon	Accuracy checked on Standard data set	DL
[148]	2019	robust approach for human activity recognition which uses the open source library *OpenPose* to extract anatomical key points from RGB images.	Camera vision based	Good accuracy on standard datasets. Good approach for independent living for elderly	OpenPose can be difficult to use	Standard dataset	RNN with LSTM

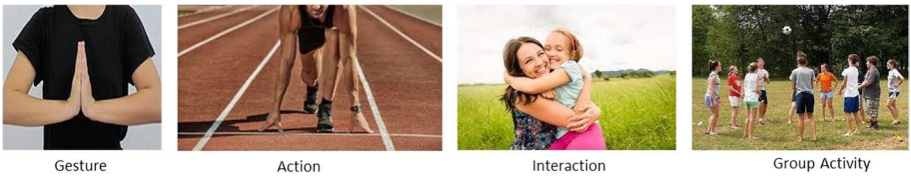

Fig. 6 Categories of Activities

activities at or near critical transportation assets. Activity recognition through Gait is the process of identifying an activity by the manner in which they walk. The identification of human activities in a video, such as a person walking, running, jumping, jogging, etc. is an important activity in video surveillance. Gupta et al. [86], contribute to the use of a Model-based approach for activity recognition with the help of the movement of the legs only.

Researchers have also worked on an efficient depth video-based HAR system that monitors the activities of elder people 24 hours a day and provides them an intelligent living space which comforts their life at home.

HAR is also influenced by culture of people. This was established in the work [139] which proposed that daily life activities, such as eating and sleeping, are deeply influenced by a person's culture, hence generating differences in the way a same activity is performed by individuals belonging to different cultures and that by taking cultural information into account we can improve the performance of systems for the automated recognition of human activities. They proposed four different solutions to the problem and used a Naive Bayes model to associate cultural information with semantic information extracted from still images. They used a dataset of images of individuals lying on the floor, sleeping on a futon and sleeping on a bed. Vision based HAR has further processing steps which will be discussed in later sections.

4 Datasets: RGB, RGB-D

There has been a lot of research based on RGB images [11, 50, 94] in the past decades. Computer vision and RGB images are closely associated with each other. RGB images give the information about the appearance of objects in the frame. The information about only the colours is a little less though as with this limited information it is a cumbersome task to partition the foreground and background having similar colours and textures. Additionally, the object appearance described by RGB images is not sturdy against common variations, such as changing illumination conditions, which hinders the usage of RGB based vision algorithms in real. While a lot of work has been going on to design sophisticated algorithms, research is parallelly going for finding a new type of representation that can better represent the scene information. Due to its complementary nature of the depth information and the visual (RGB) information, RGB-D image/video is an emerging data representation that is able to help solve fundamental problems. Meanwhile, it has been proved that combining RGB and depth information in high-level tasks (i.e., image/video classification) can dramatically improve classification accuracy [179, 180]. Table 3 shows examples of some of the RGB datasets for activity recognition whereas Table 4 lists datasets available for indoor activity recognition.

These datasets divided into two categories based on the modalities in which they are recorded such as RGB and RGB-D. Before 2010, a large number of RGB video dataset was

Table 3 RGB DATASETS

Name of the Dataset	Year	Purpose	Quality/ Format/ Source of preparation	FPS/Remarks	Action Types/Activities covered
FineGym [70]	2020	It provides temporal annotations at both action and sub-action levels with a three-level semantic hierarchy	RGB videos of gymnasts. 303 Competition records ~708 hours	High-quality videos 720P or 1080P	gymnasium videos dataset
HACS [213]	2019	source for spatiotemporal feature learning	HACS clips includes: 1.55 M 2-sec clips on 504 K videos. HACS segment includes: 140 K complete segments on 50 K videos	RGB videos of 200 actions category	504 K untrimmed videos and 1.5 M annotated clips were sampled from them. HACS Segments contains 139 K action segments densely annotated in 50 K untrimmed videos spanning 200 action categories
20BN-Something-Something Dataset V2 [158]	2017	Large collection of labeled video clips that show humans performing pre-defined basic actions with everyday objects	Quality is 100px. FPS = 12	RGB-labeled videos of sub-activities	220,847 videos, with 168,913 in the training set, 24,777 in the validation set and 27,157 in the test set. There are 174 labels
Kinetics [110]	2017	high-quality dataset for human action recognition in videos	RGB High-quality video dataset of 650,000 video clips;	Covers 400/600/700 activities lasting 10 seconds Dataset contains URLs of the videos.	500,000 video clips covering 600 human action classes with at least 600 video clips for each action class
Watch-n-Patch [199]	2015	focus on modelling human activities, comprising multiple actions in a completely unsupervised setting	RGB-D dataset	Videos capturing daily activities. Kinect sensor used for skeleton data ground truth annotations.	seven subjects perform daily activities in eight offices and five kitchens with complex background
Penn Action [96]	2013	human joint annotations for each sequence	RGB frames within 640×480	15 different actions	2326 video sequences of 15 different actions
UCF101 [173]	2012	UCF101 is one of the largest datasets of human actions	101 action categories are grouped into 25 groups containing 4–7 videos each	RGB videos	classified into 101 categories consisting of 13,320 video clips

Table 3 (continued)

Name of the Dataset	Year	Purpose	Quality/ Format/ Source of preparation	FPS/Remarks	Action Types/Activities covered
HMDB51 [169]	2011	realistic videos from various sources, including movies and web videos	Youtube, Google	RGB videos	51 action categories such as "jump", "kiss" and "laugh", with each category containing at least 101 clips, 6849 video clips
Kth [162]	2004	most standard datasets and first datasets for activity recognition, which contains six actions:	25 individuals participated as actors	RGB videos	walk, jog, run, box, hand-wave, and hand clap

Table 4 Indoor Activities Datasets

Name of the dataset	Indoor activities covered	Source of preparation	Format/Resolution/ Quality
Toyota Smarthome [53]	Consists of common daily life activities of 18 subjects;	Actors senior people of age 60–80 years old	7 Kinect v1 camera; 3 modalities: RGB, Depth and 3D skeleton; Resolution 640×480
STAIR Actions [174]	100 everyday human action categories. Each category contains around 900 to 1800 trimmed video clips.	Clips taken from Youtube and crowd-source workers	RGB videos and clips; 109,478 total videos; 100 categories
MPII Human Pose dataset [143]	25 K imagesof everyday human activities containing over 40 Kpeople with annotated body joints.	Image dataset extracted from Youtube	Every day human activities;410totalactivities;Labelled dataset;coversaseparateclassforhomeactivities
MoVi [80]	everyday actions and sports movements, and one self-chosen movement	60 female and 30 male actors with their 20 predefined activities	Captured using optical motion capture system, video cameras, inertial measurement units IMU); dataset contains 9 hours of motion capture data, 17 hours of video data from 4 different points of view (including one hand-held camera), and 6.6 hours of IMU data
HomeActionGenome [88]	Large-scale Multiview video database	Dataset prepared with actors using synchronized multi-view cameras including egocentric view.	30 hours of video; 70 daily activities; 453 classes of atomic actions; multiple modalities dataset
ActivityNet [9]	The largest benchmark for temporal activity containing 200 different types of activities and a total of 849 hours of videos collected from YouTube	Dataset relied on crowd and Amazon mechanical turk was used	203 classes of activities; average of 137 untrimmed videos per class and 1.41 activity instances per video; total of 849 hours video.
MSRDailyActivity3D [68]	There are 16 activity types: drink, eat, read a book, call cellphone, write on paper, use a laptop, use a vacuum cleaner, cheer up, sit still, toss paper, play games, lay down on the sofa, walk, play guitar, stand up, sit down	Daily activities captured by Kinect device.	Total activity samples is 320
GTEA (Georgia Tech Egocentric Activity) [192]	contains seven types of daily activities such as making sandwiches, tea, or coffee,	performed by four different people, thus a total of 28 videos.	Seven activities by 4 people 28total videos; 20 fine-grained instances for every video

Table 4 (continued)

Name of the dataset	Indoor activities covered	Source of preparation	Format/Resolution/ Quality
CAD-120 [36]	sequences of humans performing activities in the kitchen, living room and office environment etc.	Depth videos recorded using Kinect sensor; 4 people in different environments	RGB-D videos
Florence3D [140]	activities like wave, drink from a bottle, answer phone, clap, tight lace, sit down, stand up, read watch, bow	10 subjects performed all activities 2/3 times; Kinect was used	Total 215 activity samples; depth dataset
UT-Kinect (UTKinect-Action3D Dataset) [184]	activities like walk, sit down, stand up, pick up, carry, throw, push, pull, wave hands, clap hands.	10 subjects performed 10 activities	3 channels used: RGB, depth and skeleton joint locations; 30fps;
PKU-MMD [130]	Number of activities covered in 2 phases	51 actions performed by 66 subjects using Kinect V2 sensor.	4 modalities-RGB,Depth, Infrared radiation and skeleton; videos containing 30fps; dataset contains 2 phases, phases #1 contains 51 action categories, performed by 66 distinct subjects in 3 camera views

available [40]. After the advancement of low-cost depth sensors, e.g., Microsoft Kinect, there has been a drastic increase in the 3D dataset, and multi-modal videos dataset [13, 119, 132]. Due to low cost and lightweight sensors datasets are recorded with multiple modalities such as depth frames, accelerometer, IR sensors frames, acoustical data, and skeleton data information. The RGB-D datasets have multiple modalities which reduces the loss of information in 3D videos as compared to traditional RGB datasets at the cost of increased complexities. Table 3 shows a few of the existing datasets for activity recognition including examples of multi-modal datasets as well.

The challenges and concerns which might exist for performing HAR on such complex datasets are- occlusion, cluttered background, localization of activity, noise removal, etc. The use of depth sensor like Microsoft Kinect for creating RGB-D datasets may also pose practical problems like sensor placement and orientation. At the beginning, Kinect was as an Xbox accessory, enabling players to interact with the Xbox 360 through body language or voice instead of the usage of an intermediary device, such as a controller. Later on, due to its capability of providing accurate depth information with relatively low cost, the usage of Kinect goes beyond gaming, and is extended to the computer vision field. This device equipped with intelligent algorithms is contributing to various applications, such as 3D-simultaneous localization and mapping (SLAM) [89, 117] people tracking [151], object recognition [30] and human activity analysis [42, 128] etc.

5 Approaches of HAR

Activity recognition systems classify and recognise various activities for which they have been trained. Since, the data and variety of data is always better for such systems, the features from this data are again crucial and significant part of the HAR process. This leads to two types of approaches for the feature extraction- handcrafted features (machine learning), automated feature learning (deep learning) method. Traditionally, the researchers used the manual extraction of features from data (handcrafted features) [31, 116] for training the machine learning models. However, with the development of deep neural networks [112], models tend to automatically learn features. With great success of CNNs on image classification tasks, CNNs were applied to recognize human activities in videos. Well known approaches for action recognition in RGB videos based on CNNs are presented in [107, 123, 171, 194, 196]. With the introduction of cost inexpensive RGB-D sensors, researchers shift their attention, towards other vision cues like depth and skeletal data along with RGB data. One of the advantages of depth data and skeletal data, as compared to traditional RGB data is that they are less sensitive to changes in lighting conditions. Furthermore, the availability of well-known and diverse RGB-D datasets like MSR Action 3D [58], UTD-MHAD [43], Berkeley MHAD [150], CAD-60 [178], SBU Kinect interaction [150], and many more, encouraged extensive research for human activity recognition using multiple vision cues. However, the two approaches are efficient in their own respect. Table 5 lists a few of the recent works using the two above-mentioned approaches.

The results of HAR models for activity recognition also vary with how the human body is represented. Based on the category of representation of a Human body, its representation can be classified into 2D and 3D representation methods. 2D methods take into account spatial information, contour, shape, and skeletal information. 3D representation on the other hand talks of skeletal joints in terms of depth together with other characteristics such as spatial and temporal components related to the frame.

Table 5 Handcrafted and automated feature in HAR

Ref.	Year	Approach	Objective	Dataset used	Performance measures
[159]	2021	Automatically learned features	A neural network (NN) based approach for the classification and evaluation of human activities has been explored together with long short-term memory (LSTM)	KARD, CAD-60, MSR DAILY ACTIVITY 3D	Precision, Recall on KARD: 100% Precision, recall on CAD-60:100% Accuracy on MSR daily 95.6%
[8]	2021	Handcrafted features	Recognition of significant joints using features from spatial and temporal event feeds.	Real-time event Video dataset	10–15% memory usage over 532 MB of digitised real-time event data with 0.5431 sec processing time
[137]	2021	Automatically learned features	The work proposed a novel deep-learning framework to solve the HAR effect on overall accuracy. The framework is a location-based CNN-LSTM hybrid model.	DHA dataset with CNN-LSTM approach shows best accuracy	Acc. Of 96.75%
[172]	2021	Automatically learned features	a deep bottleneck multimodal feature fusion (D-BMFF) framework that fused three different modalities of RGB, RGB-D(depth) and 3D coordinates information for activity classification	Four RGB-D datasets: UT Kinect, CAD-60, Florence 3D, and SBU Interaction	The ARR on UT Kinect, CAD-60, Florence 3D, SBU Interaction are 99%, 98.50%, 98.10%, and 97.75% respectively
[209]	2021	Automatically learned features	approach for human activity recognition using ensemble learning of multiple convolutional neural network (CNN) models		Ensemble of CNN models gives accuracy of 94%
[26]	2021	Automatically learned features	deep learning-based method for human activity recognition problem. The method uses convolutional neural networks to automatically extract features from raw sensor data and classify six basic human activities	Diabetes dataset	Different activities covered with approximately 90% accuracy using CNN,Random forest, SVM

Table 5 (continued)

Ref.	Year	Approach	Objective	Dataset used	Performance measures
[135]	2021	Automatically learned features	a deep learning architecture that leverages the feature extraction capability of the convolutional neural networks and the construction of the temporal sequences of recurrent neural networks to improve existing classification results		Accuracy increased from 30% to 35% due to transfer learning
[57]	2021	Automatically learned features, handcrafted features	framework to extract handcrafted high-level motion features and in-depth features by CNN in parallel to recognize human action. SIFT is used as a handcrafted feature to encode high-level motion features from the maximum number of input video frames. The combination of deep and handcrafted features preserves more extended temporal information from entire video frames present in action video with minimal computational power	UCF and KTH	89.5%; CNN combined with SIFT gives a lesser vector dimension for the model making it computationally efficient.

Table 5 (continued)

Ref.	Year	Approach	Objective	Dataset used	Performance measures
[92]	2020	Automatically learned features	propose a deep learning multi-channel architecture using a combination of convolutional neural network (CNN) and Bidirectional long short-term memory (BLSTM). The advantage of this model is that the CNN layers perform direct mapping and abstract representation of raw sensor inputs for feature extraction at different resolutions. The BLSTM layer takes full advantage of the forward and backward sequences to improve the extracted features for activity recognition significantly.	WISDM dataset	For 5 sec: 98.6±0.682 For 10 sec: 99.1±0.455 For 15 sec: 92.6±0.431 For 20 sec: 82.7±0.513
[144]	2020	Automatically learned features, Hand-crafted features	aims in capturing the motion information of the whole video by producing a dynamic image corresponding to the input video; use two parallel ResNet-101 architectures to produce the dynamic images for the RGB video and depth video separately	MSR Action 3D Dataset and proposed dataset on which many existing algorithms were tested.	MSR Action 3D: approx. 90%; proposed dataset had 54.94% with processing and 50.12% without processing.
[59]	2020	Automatically learned features	a deep view-invariant human action recognition framework, which is a novel integration of two important action cues: motion and shape temporal dynamics (STD)	NUCLA multi-view dataset, UWA3D-II Activity dataset and NTU RGB-D Activity dataset	NUCLA- 87.3% UWA3D-II –85.2% NTU RGBD- 79.4%
[23]	2020	Automatically learned features, hand-crafted features	combines a traditional classifier based handcrafted feature extractor and a deep learning-based method in order to replace the artisanal feature extraction method with a new one.	MSR Daily Activity 3D Dataset	99.92%: pretrained CNN+ SVM 98.77% precision; 99.79- recall; F-measure: 99.28

Table 5 (continued)

Ref.	Year	Approach	Objective	Dataset used	Performance measures
[10]	2020	Automatically learned features	a Lightweight Deep Learning Model for HAR requiring less computational power, making it suitable to be deployed on edge devices. The performance of proposed model is tested on the participant's six daily activities data	RNN-LSTM trained on WISDM	RNN-LSTM- 99% Accuracy for jogging, walking; min 81% accuracy for upstair activity
[75]	2020	Handcrafted features	approach for activity recognition combining RGB data and skeleton analysis using Kinect sensors	CAD-60, CAD-120, OAD Dataset internally aquired	CAD-60: precision,recall Skeleton 95.0 95.0; RGB (20 sectors) 92.5 89.4; Score-level fusion 98.8 98.3 CAD-120: precision,recall Skeleton 77.6 73.;1 RGB (20 sectors) 61.1 59.3 Score-level fusion 85.4 83.3 OAD: precision,recall Skeleton 80.6 80.5; RGB (20 sectors) 85.8 85.9; Score-level fusion 90.6 90.4

Table 5 (continued)

Ref.	Year	Approach	Objective	Dataset used	Performance measures
[183]	2019	Automatically learned features, hand-crafted features	a novel approach to recognize human actions by considering both deep spatial features and handcrafted spatiotemporal features. Firstly, the deep spatial features by employing a state-of-the-art deep convolutional network, namely Inception-Resnet-v2 are extracted. Secondly, they introduced a novel handcrafted feature descriptor, namely Weber's law based Volume Local Gradient Ternary Pattern (WVLGTP), which brings out the spatiotemporal features. It also considers the shape information by using gradient operation	KTH, UT Interaction, Hollywood2, UCF101, UT INTERACTION	On KTH dataset, proposed method (Inception-resnet-v2 with WVLGTP) shows 96.5% accuracy; On UT interaction, it is 97.6%; ON Hollywood2, accuracy is 70.3%; On UCF-101, it is 94.9%;
[191]	2019	Automatically learned features	designs a smartphone inertial accelerometer-based architecture for HAR. When the participants perform typical daily activities, the smartphone collects the sensory data sequence, extracts the high-efficiency features from the original data, and then obtains the user's physical behaviour data through multiple three-axis accelerometers	UCI HAR dataset, Pamap2 dataset	UCI HAR dataset: CNN, LSTM, BLSTM, MLP, SVM Average acc 0.9100 0.8586 0.8952 0.8272 0.8407 Respectively. Pamap2 dataset: CNN, LSTM, BLSTM, MLP, SVM average acc 0.9321 0.8914 0.8941 0.8683 0.9050 respectively.
[46]	2019	Automatically learned features, hand-crafted features	A fusion of handcrafted features with automatically learned features by a deep algorithm for HAR using smartphones	A public dataset and proposed dataset	On proposed algorithm, Maximum Full a Posterior (MFAP) accuracy is 98.85%

Table 5 (continued)

Ref.	Year	Approach	Objective	Dataset used	Performance measures
[45]	2018	Handcrafted features	An argument is proposed that feature embedding from deep neural networks may convey complementary information and propose a novel knowledge distilling strategy to improve its performance. More specifically, an efficient shallow network, i.e., single-layer feedforward neural network (SLFN), with handcrafted features is utilized to assist a deep long short-term memory (LSTM) network.	Proposed dataset readied as per the protocols.	97.7%
[138]	2018	Automatically learned features, hand crafted features	a new deep learning network for action recognition that integrates quaternion spatial-temporal convolutional neural network (QST-CNN) and Long Short-Term Memory network (LSTM), called QST-CNN-LSTM	Weizmann, UCF sports, UCF11	On Weizmann: QST-CNN-LSTM −96.34% recognition accuracy, 3D CNN, Gray-CNN, and 3Channel-CNN were 90.12%, 86.46%, and 76.00%, respectively
[14]	2018	Automatically learned features	a robust position-independent HAR system for smartphone using a deep CNN model	New dataset	98%
[18]	2017	Automatically learned features, handcrafted features	Uses RGB-D data for recognition of the action using multimodal fusion by using scene flow as early fusion and integrating the modalities' data in a late fusion fashion	Montalbano II, MSR Daily Activity 3D	DT and MMDT accuracy on MSR Daily Activity 3D is 63.125%, 78.13%; DT and MMDT accuracy on Montalbano is 83.5% and 85.66%
[136]	2017	Handcrafted features	The method tries state-of-art classifiers and explains the role of feature selection for activity recognition in smartphones	State-of-the-art classifiers: Bayes classifier, kNN, MLP, SVM, MLM and MLM-NN	MLM and SVM achieved accuracy of more than 99.2% in the original data set and 98.1% using new feature selection method

The skeleton is a high-level presentation that can be used to describe human activity in a very precise way and is adapted to the challenge of activity analysis, pose estimation, and action recognition. However, skeleton data include the coordinates of the human body's key joints over time. This is an important factor for the motion presentation of each action. Body pose is an important indicator of human actions. Human body pose is a type of high-level semantic feature that has shown effect in recognizing human actions with discriminative geometric relations with body joints. Recent researches focus on either 2D or 3D modeling of these body joints. Most of the latest research in the field of automated HAR in fact based on skeleton data and uses depth devices such as Kinect to obtain three-dimensional (3D) skeleton information directly from the camera. Although these researches achieve high accuracy but are strictly device dependent and cannot be used for videos other than from specific cameras.

The 2D representation approach is effective for applications where precise pose recovery is not needed or possible [154]. The low-resolution image or single viewpoint can also be a reason for 2D representation methods. Example surveillance camera positioning. 3D approaches are suitable for a scenario involving a high level of discrimination between various unconstrained and complex human movements (Table 6).

6 Challenges in indoor HAR

Human activity recognition is an active and pervasive field for researchers not only from the areas of Computer Science, Electrical Engineering, and Informatics but also from many interdisciplinary backgrounds. This is because this field involves many challenges. From capturing data to the recognition process, all steps involve a thoughtful procedure that can be helpful in devising an appropriate way of recognizing the targeted activities. The environment for regular activities can be constrained and noisy.

The hardware involved has to be carefully thought out. One has to take into consideration real-world hardware implementation, efficient and resource-constraint computation, novel physics-based sensing, and socio-economical and human factor-related challenges.

The activities for which the model is to be built, also help in deciding the approach ahead. The indoor activities performed in controlled and static background conditions may not require a great deal of complex pre-processing steps whereas localization and denoising become an integral part of the process involving activities performed with complex backgrounds and multiple objects in a single frame. Choosing features is again a challenge for building a good model. Traditionally, machine learning used handcrafted features but with the popular and successful implementation of convolution neural network models, the deep learning approach has gained momentum and is a preferred choice for researchers. Thus, broadly talking of the levels of challenges in the pipeline for HAR we can categorize them as Application-level challenges, Data acquisition-level challenges, Approach level challenges, Algorithm level challenges, and Implementation level challenges. These are represented in Fig. 7 below.

6.1 Application-level challenges

The application area of indoor HAR is a big factor to study the challenges involved. The general application areas of HAR are surveillance, elderly care [176, 210], assisted living [185], smart homes, abnormal activity recognition [115, 131, 215], sports [164], etc. The

Table 6 2D and 3D representation approach

Ref.	Year	2D/3D	Objective	Model Used/Remarks
[104]	2020	2D,3D	Better presentation of the human body using different features and the exploitation of the RNN structure for activities.	Training using human skeleton features; Recognition using GRU + RNN
[62]	2019	2D	2d representation of the human body using skeleton key points and contour key points for human shape information	depth differential silhouettes (DDS) between two consecutive frames, and it is represented by the histogram of oriented gradients (HOG) format; multi-fused features + Accumulated HMM
[79]	2019	2D	2D skeletal data is extracted from videos obtained from a normal camera for activity recognition. Uses OpenPose library to find appearance and motion features using 2D positions of human skeletal joints	Comparisons of KNN, SVM, LDA, NB, and BPNN used for fall, walk, and sitting activities and comparison with state-of-the-art for the proposed work with normalized skeleton data
[121]	2018	2D	an approach based on a deep neural network architecture for 2D pose-based action recognition tasks	The use of pose features and global features for multi-modal action recognition made the system perform better than the P-CNN and JDD which are recent pose-based approaches.
[200]	2018	3D,2D	first method to capture the 3D total motion of a target person from a monocular view input	POFs are predicted by a Fully Convolutional Network, along with the joint confidence maps
[52]	2018	2D,3D	Motion information gathered from skeleton joint sequences and use of ConvNets based approach for activity recognition by combining multiple vision cues	Combination of multiple cues by means of decision level fusion.
[103]	2017	3D	multi-fused features for online human activity recognition (HAR) system that recognizes human activities from continuous sequences of depth map	Combining spatio-temporal features and depth silhouettes give a better idea of the activity. Forward spotting scheme was also proposed to differentiate between activity and non-activity.
[102]	2016	3D	spatiotemporal features approach to detect, track, and recognize human silhouettes using a sequence of RGB-D images	Use of spatiotemporal data from depth images is a good feature that increased the accuracy as it can be used for storing data in case of sequential activities.
[126]	2015	3D	A deep structured model, which decomposes an activity instance into temporal parts using the CNN	The temporal feature helps in considering large variations within activities and sub-segments of the activities are taken into account for better results of the deep learning model; CNN with back propogation

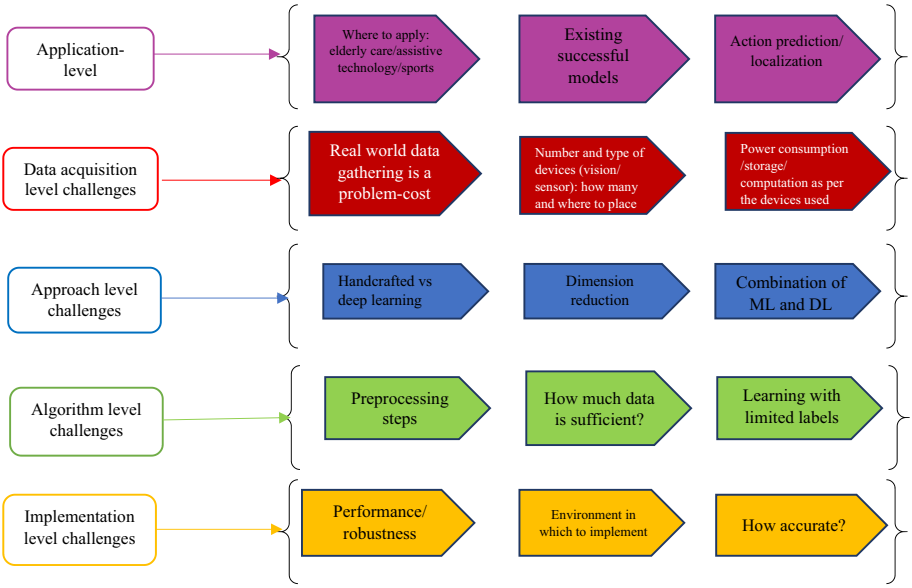

Fig. 7 Various levels of challenges and factors involved in Indoor HAR

challenges vary according to the area of application. Application of the model will help in deciding the steps ahead for building a system that considers the challenges which exist in that research problem.

6.2 Data acquisition level challenges

As previously discussed, the data for indoor HAR which can be a video or an image from the video can be either captured using vision-based methods or sensor-based methods. Some studies have combined the image/video under the sensors category for the sake of simplicity. This survey has considered vision-based input but also discussed current research in both the sensor-based and vision-based input fields.

6.2.1 Real-world data gathering

The data captured from whichever medium may have noise due to the environment from which the data has been extracted. For example, speech data from a kitchen activity will have background noise. Similarly, the images or video captured from cameras placed indoors in the living room can have noise in the form of dust particles, low visibility [122, 155] etc. Due to this noise the accuracy of the model may suffer.

6.2.2 Number and type of devices single camera or multiple cameras

The other problem is the choice of device to capture the data. Data can be captured from a static camera installed in the region of the activity or it can be collected using types of sensors (wearable, ambient) as discussed in section 2. In both cases the number and location

of the sensor/camera matter. The data may lose its worth if it loses the critical features which are important for recognizing a particular activity. The capturing devices should be thoughtfully placed. The application for which it is being used will also matter together with the choice of sensor. Like, if the monitoring of the elderly is to be done using a wearable sensor, then one sensor is sufficient on the wrist. On the other hand, if the camera-based sensor is to be used then we may need two cameras or even more to detect and capture the whole activity being done.

6.2.3 Power consumption/storage/computation/Privacy

With the increase in application areas of activity monitoring, CCTVs have gained popularity in various segments involving AAL [54] elderly care [165]. But the challenge is the increasing demand of power, storage and computation which are essential pillars to the success of a model for activity recognition. These factors cannot be overlooked and thus reduction in terms of these parameters play as a gamechanger for the model.

Apart from storage and power consumption, privacy of data and person are also very important. The elderly person may feel uncomfortable being monitored by someone for his daily activities in case of a camera. While using wearable sensors, other issues related to networking may cause a concern.

6.3 Approach level challenges

The recognition and analysis by a model require knowing in depth of the features which can prove efficient. Knowledge of feature dimension and interpretability is also a key aspect to create a powerful yet computationally light model.

6.3.1 Handcrafted vs. deep learning

The impressive classification performance of the deep learning methods is making it popular among the researchers. However, it is known that they typically require a large training sample to achieve that accuracy. Meanwhile, handcrafted (HC) features have been implemented for decades and still serve as a powerful tool when combined with machine learning classifiers. Researchers are still exploring the best technique to suit their problem [16, 127]. Table 5 listed some of the latest research works using either or both of these approaches.

6.3.2 Dimension reduction

Another important challenge is to reduce the size of the feature so that the computation is easy to process. Also, a non-contributing feature should not be part of the process. Researchers have been working on this aspect for a better solution to the problem. Table 7 lists some of the recent works in the dimensionality reduction area for deep learning and machine learning techniques.

6.3.3 Combination of automatically-calculated and handcrafted features

Researchers have been trying to combine handcrafted features (machine learning) and automatically-calculated (deep learning) features to design such systems which can be accurate and robust. Table 5 lists a few of the recent works which have combined both types of

Table 7 Few of the recent works for dimensionality reduction

Ref.	Year	Area	Objective	Method used/Remarks
[161]	2021	Deep learning	proposes a hybrid approach to analyse and recognize human activity on the same dataset using deep learning methods on a cloud-based platform and principal component analysis is applied to the dataset to get the most important features.	561 features to 48 using 50 PCA components; CNN used further
[142]	2021	Sensor data	proposed a dimensionality reduction technique called fast feature dimensionality reduction technique (FFDRT)[reduces the number of features to 561 to 66]	Input features: 6000×561; output: 6000×66; classification algorithm used: KNN, random forest, deep learning
[201]	2019	CNN	a network intrusion detection model based on a convolutional neural network-IDS (CNN-IDS). Redundant and irrelevant features in the network traffic data are first removed using different dimensionality reduction methods. Features of the dimensionality reduction data are automatically extracted using CNN, and more effective information for identifying intrusion is extracted by supervised learning.	CNN-IDS, DNN, RNN
[156]	2017	Deep learning, hand-crafted features	investigates a particular approach to combine hand-crafted features and deep learning to (i) achieve an early fusion of off-the-shelf handcrafted global image features and (ii) reduce the overall number of dimensions to combine both worlds. This method allows for fast image retrieval in domains, where training data is sparse.	Reduction in processing time; handcrafted and deep learning;
[66]	2016	Machine learning	to deal with HAR modeling involving a significant number of variables in order to identify relevant parameters from data and thus maximize the classification accuracy while minimizing the number of features using data mining techniques	Dimensionality reduction models: Pristine, PCA with KNN, C5.0 as recognition models
[44]	2016	Sensor data	introduce the framework of a manifold elastic net that encodes the local geometry to find an aligned coordinate system for data representation.	PCA for dimensionality reduction
[195]	2016	Deep learning	to investigate the dimensionality reduction ability of the auto-encoder, and see if it has some kind of good property that might accumulate when being stacked and thus contribute to the success of deep learning.	Auto-encoders for dimensionality reduction
[193]	2014	Neural network	a dimensionality reduction method by manifold learning, which iteratively explores data relations and uses the relation to pursue the manifold structure.	Autoencoder used for dimensionality reduction; deep autoencoders used for complex datasets

Table 8 Summary of Proposed Taxonomy in terms of Components

Components of Vision-based indoor HAR	Remarks/Challenges
Types of Input • Camera • Depth (camera) sensor	Pros: No dependence on wearing/ placement of sensors; No battery issues. Cons: Camera cost/ Installation angles/ number of cameras/ sensors to be deployed/ application for which used
Body Representation • 2D • 3D	Human body 2D representation low computation and storage requirement; Relies on edges and shape; 3D considers depth; requires depth sensors like Kinect
Modalities • 2D • 3D • Skeleton based	*2D* relies of visual RGB content only; *3D* has a depth channel; gives better results; *Skeleton-based attaches* information of standard key-points of human body; Pose estimation and activity becomes clear.
Solutions • Handcrafted features (Global feature-based, local feature-based) Approaches: Space-time trajectory, Space-time features, Shape-based, motion-based, LBP, fuzzy logic, etc. • Learned feature representation/ Deep learning based > Spatio-temporal network > Multiple stream network > Deep Generative Network > Temporal coherency network	*Handcrafted Solutions* Popular methods: HMM, BOW, HOG, Bayesian network, etc. Pros: Works well for lesser data; Cons: challenges like lighting, and occlusion may impact the system's accuracy to greater lengths. *Deep Learning Solutions:* Popular methods: CNN, LSTM, CNN-3D, RNN,etc. Pros: Appropriate for sequential processing; temporal data with spatial data can help in increasing the accuracy; Cons: requires large amounts of data. Costly set up required
Challenges • Application level • Data acquisition level • Approach level • Algorithm level • Implementation level	*Where* to apply; *What* application area is involved; *How* to acquire the data; *Which* approach is suitable; *Which* algorithm is to be chosen; *Robustness, performance measures, etc.*

features for HAR models. This is a challenging task to choose the appropriate techniques and compare them with other models.

6.4 Algorithm-level challenges

This challenge talks about the concerns like pre-processing steps to be included, the data volume to be taken, and learning with limited labels. These concerns and issues are correlated and overlap with the issues discussed above. For a better HAR performance, data pre-processing is a crucial step in machine learning and deep learning algorithm. Data

pre-processing includes data cleaning, normalization, transformation, feature extraction, and selection [111].

An important issue in HAR is whether to use training data from a general population (subject-independent), or personalized training data from the target user (subject-dependent). Past research has shown better results with personalized data but a collection of end-user data for training is not a practical option. Thus, the subject-independent approach is more common. [164] introduce a novel approach that uses nearest-neighbor similarity to identify examples from a subject-independent training set that are most similar to sample data obtained from the target user and uses these examples to generate a personalized model for the user.

6.5 Implementation-level challenges

Implementation of the model finally gives us an idea of how the system behaves in real-time. This arises because of the challenges like low lighting conditions, occlusions, etc. Overall, the requirement is for a robust and reliable system. These challenges also include network issues which may cause real-time glitches in HAR. The accuracy and time to execute, both are important aspects of a real-time system. If the system takes too long to process then the lag decreases the real-time effect thus making the system inefficient. If the system accuracy is less but the speed of processing is good then the system is unreliable. So, both concerns should be addressed equally.

7 Conclusion and future work

In this survey, we have reviewed recent trends in the field of indoor human activity recognition. We have proposed a taxonomy for indoor HAR. The review is conducted on the state-of-the-art HAR input methodology, approaches, datasets, and human body representation methods. The datasets which have been used in various recent research have alsso been listed. The survey lists various challenges associated with indoor HAR at various levels. Indoor-HAR can provide smart solutions for elderly care, patient care systems, physical fitness systems, abnormal human behavior detection, and analysis, etc. We can monitor our indoor activities to observe the change in activity patterns before the onset of any disease or major illness. These can help in situations where these changes go unnoticed because we tend to ignore certain changes in the body. Such systems can generate data and help us observe physical changes. The cases of child abuse during the pandemic had increased. Such systems can help in such situations too for detecting any abnormal behavior, fall detection, domestic violence, child abuse, etc. Table 8 summarises the survey in terms of the individual components of the proposed taxonomy. Indoor-HAR is a vast area with many key and challenging applications.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. "(4) (PDF) Human Activity Recognition from Video: modeling, feature selection and classification architecture." (n.d.) https://www.researchgate.net/publication/237448747_Human_Activity_Recognition_from_Video_modeling_feature_selection_and_classification_architecture. Accessed 19 Aug 2021

2. "(4) (PDF) Real time detection, tracking and recognition of medication intake." (n.d.) https://www.researchgate.net/publication/242772663_Real_time_detection_tracking_and_recognition_of_medication_intake. Accessed 19 Aug 2021
3. "(8) (PDF) Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map." (n.d.) https://www.researchgate.net/publication/326175323_Dense_RGB-D_Map-Based_Human_Tracking_and_Activity_Recognition_using_Skin_Joints_Features_and_Self-Organizing_Map. Accessed 28 Aug 2021
4. "(8) Development of a life logging system via depth imaging-based human activity recognition for smart homes | Request PDF." (n.d.) https://www.researchgate.net/publication/292224847_Development_of_a_life_logging_system_via_depth_imaging-based_human_activity_recognition_for_smart_homes Accessed 28 Aug 2021
5. "(8) Human activity recognition using the labeled depth body parts information of depth silhouettes | Request PDF." (n.d.) https://www.researchgate.net/publication/329877194_Human_activity_recognition_using_the_labeled_depth_body_parts_information_of_depth_silhouettes. Accessed 28 Aug 2021
6. Abdel-Basset M, Chang V, Hawash H, Chakrabortty RK, Ryan M (2021) Deep learning approaches for human-centered IoT applications in smart indoor environments: a contemporary survey. Ann Oper Res 2021:1–49. https://doi.org/10.1007/S10479-021-04164-3
7. Abdelgawwad A, Mallofre AC, Patzold M (2021) A Trajectory-Driven 3D Channel Model for Human Activity Recognition. IEEE Access 9:103393–103406. https://doi.org/10.1109/ACCESS.2021.3098951
8. Abdul Lateef Haroon PS, Premachand DR (2021) Human Activity Recognition using Machine Learning Approach. J Robot Control (JRC) 2(5):395–399. https://doi.org/10.18196/JRC.25113
9. "Activity Net." (n.d.) http://activity-net.org/. Accessed 23 Nov 2021
10. Agarwal P, Alam M (2020) A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. Procedia Comput Sci 167:2364–2373. https://doi.org/10.1016/J.PROCS.2020.03.289
11. Aggarwal JK, Cai Q (1999) Human Motion Analysis: A Review. Comput Vis Image Underst 73(3):428–440. https://doi.org/10.1006/CVIU.1998.0744
12. Aggarwal JK, Ryoo MS (2011) Human activity analysis. ACM Comput Surv (CSUR) 43(3):43. https://doi.org/10.1145/1922649.1922653
13. Aggarwal JK, Xia L (2014) Human activity recognition from 3D data: A review. Pattern Recogn Lett 48:70–80. https://doi.org/10.1016/J.PATREC.2014.04.011
14. Almaslukh B, Artoli AM, Al-Muhtadi J (2018) A robust deep learning approach for position-independent smartphone-based human activity recognition. Sensors 18(11):1–17. https://doi.org/10.3390/s18113726
15. Amirbandi EJ, Shamsipour G (2016) Exploring methods and systems for vision-based human activity recognition. 2016 1st conference on swarm intelligence and evolutionary computation (CSIEC), Bam, Iran, 2016, pp 160–164. https://doi.org/10.1109/CSIEC.2016.7482122
16. Antipov G, Berrani SA, Ruchaud N, Dugelay JL (Oct. 2015) Learned vs hand-crafted features for pedestrian gender recognition. MM 2015 - Proceedings of the 2015 ACM Multimedia Conference, pp. 1263–1266. https://doi.org/10.1145/2733373.2806332
17. Arifoglu D, Bouchachia A (2019) Detection of abnormal behaviour for dementia sufferers using convolutional neural networks. Artificial Intelligence in Medicine 94:88–95. https://doi.org/10.1016/J.ARTMED.2019.01.005
18. Asadi-Aghbolaghi M, Bertiche H, Roig V, Kasaei S, Escalera S (2017) Action Recognition from RGB-D data: comparison and fusion of spatio-temporal handcrafted features and deep strategies. pp. 3179–3188
19. Ashraf I, Zikria YB, Hur S, Bashir AK, Alhussain T, Park Y (2021) Localizing pedestrians in indoor environments using magnetic field data with term frequency paradigm and deep neural networks. Int J Mach Learn Cybern 12:3203–3219. https://doi.org/10.1007/S13042-021-01279-8
20. Ayase R, Higashi T, Takayama S, Sagawa S, Ashida N (2008) A method for supporting at-home fitness exercise guidance and at-home nursing care for the elders, Video-based simple measurement system. *2008 10th IEEE Intl. Conf. on e-Health Networking, Applications and Service, HEALTHCOM 2008*, pp. 182–186. https://doi.org/10.1109/HEALTH.2008.4600133
21. Babiker M, Khalifa OO, Htike KK, Hassan A, Zaharadeen M (Mar. 2018) Automated daily human activity recognition for video surveillance using neural network. *2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017*, vol. 2017-November, pp. 1–5. https://doi.org/10.1109/ICSIMA.2017.8312024
22. Bao L, Intille SS (2004) Activity Recognition from User-Annotated Acceleration Data. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 3001, pp. 1–17. https://doi.org/10.1007/978-3-540-24646-6_1

23. Basly H, Ouarda W, Sayadi FE, Ouni B, Alimi AM (Jun. 2020) CNN-SVM Learning Approach Based Human Activity Recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12119 LNCS, pp. 271–281. https://doi.org/10.1007/978-3-030-51935-3_29.
24. Beddiar DR, Nini B, Sabokrou M, Hadid A (2020) Vision-based human activity recognition: a survey. Multimed Tools Appl 79(41–42):30509–30555. https://doi.org/10.1007/S11042-020-09004-3
25. Ben-Arie J, Wang Z, Pandit P, Rajaram S (2002) Human activity recognition using multidimensional indexing. IEEE Trans Pattern Anal Mach Intell 24(8):1091–1104. https://doi.org/10.1109/TPAMI.2002.1023805
26. Bhat O, Khan DA (2021) Evaluation of deep learning model for human activity recognition. Evol Syst 1:1–10. https://doi.org/10.1007/S12530-021-09373-6
27. Bhola G, Kathuria A, Kumar D, Das C (May 2020) Real-time Pedestrian Tracking based on Deep Features. *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, pp. 1101–1106, https://doi.org/10.1109/ICICCS48265.2020.9121061
28. Bibbo L (n.d.) AN INTEGRATED SYSTEM FOR INDOOR PEOPLE LOCALIZATION, TRACKING, AND MONITORING Localization and tracking of people in indoor environment View project. [Online]. Available: www.scientific-publications.net. Accessed Oct. 17, 2021
29. Blank M, Gorelick L, Shechtman E, Irani M, Basri R (n.d.) Actions as Space-Time Shapes
30. Bo L, Ren X, Fox D (2013). Unsupervised feature learning for RGB-D based object recognition. Experimental Robotics. 88. https://doi.org/10.1007/978-3-319-00065-7_27
31. Bobick AF, Davis JW (2001) The recognition of human movement using temporal templates. IEEE Trans Pattern Anal Mach Intell 23(3):257–267. https://doi.org/10.1109/34.910878
32. Bocus MJ et al (Oct. 2021) OPERAnet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision-based Sensors. [Online]. Available: https://arxiv.org/abs/2110.04239v1. Accessed 17 Oct 2021
33. Brand M, Oliver N, Pentland A (1997) Coupled hidden Markov models for complex action recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 994–999, https://doi.org/10.1109/CVPR.1997.609450
34. Bux A, Angelov P, Habib Z (2017) Vision Based Human Activity Recognition: A Review. Adv Intell Syst Comput 513:341–371. https://doi.org/10.1007/978-3-319-46562-3_23
35. Buzzelli M, Albé A, Ciocca G (2020) A Vision-Based System for Monitoring Elderly People at Home. Appl Sci 10:374. https://doi.org/10.3390/app10010374
36. "CAD-120 Dataset | Papers With Code." (n.d.) https://paperswithcode.com/dataset/cad-120. Accessed 23 Nov 2021
37. Capela NA, Lemaire ED, Baddour N (2015) Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able bodied, Elderly, and Stroke Patients. PLoS One 10(4):e0124414. https://doi.org/10.1371/JOURNAL.PONE.0124414
38. Cardone G, Cirri A, Corradi A, Foschini L, Ianniello R, Montanari R (2014) Crowdsensing in Urban areas for city-scale mass gathering management: Geofencing and activity recognition. IEEE Sensors J 14(12):4185–4195. https://doi.org/10.1109/JSEN.2014.2344023
39. Casale P, Pujol O, Radeva P (2011) Human activity recognition from accelerometer data using a wearable device. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6669 LNCS, pp. 289–296. https://doi.org/10.1007/978-3-642-21257-4_36
40. Chaquet JM, Carmona EJ, Fernández-Caballero A (2013) A survey of video datasets for human action and activity recognition. Comput Vis Image Underst 117(6):633–659. https://doi.org/10.1016/J.CVIU.2013.01.013
41. Chen Y, Shen C (2017) Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access 5:3095–3110. https://doi.org/10.1109/ACCESS.2017.2676168
42. Chen L, Wei H, Ferryman J (2013) A survey of human motion analysis using depth imagery. Pattern Recogn Lett 34(15):1995–2006. https://doi.org/10.1016/J.PATREC.2013.02.006
43. Chen C, Jafari R, Kehtarnavaz N (Dec. 2015) UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. *Proceedings - International Conference on Image Processing, ICIP*, vol. 2015-December, pp. 168–172. https://doi.org/10.1109/ICIP.2015.7350781
44. Chen YL, Wu X, Li T, Cheng J, Ou Y, Xu M (2016) Dimensionality reduction of data sequences for human activity recognition. Neurocomputing 210:294–302. https://doi.org/10.1016/J.NEUCOM.2015.11.126
45. Chen Z, Zhang L, Cao Z, Guo J (2018) Distilling the Knowledge from Handcrafted Features for Human Activity Recognition. IEEE Trans Industr Inform 14(10):4334–4342. https://doi.org/10.1109/TII.2018.2789925

46. Chen Z, Xiang S, Ding J, Li X (2020) Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori. IEEE Trans Instrum Meas 69(7):3992–4001. https://doi.org/10.1109/TIM.2019.2945467
47. Cheng G, Wan Y, Saudagar AN, Namuduri K, Buckles BP (Jan. 2015) Advances in Human Action Recognition: A Survey. [Onlisssne]. Available: https://arxiv.org/abs/1501.05964v1. Accessed 19 Aug 2021
48. Cheng X, He M, Duan W (Apr. 2018) Machine vision based physical fitness measurement with human posture recognition and skeletal data smoothing. *Proceedings of the 2017 International Conference on Orange Technologies, ICOT 2017*, vol. 2018-January, pp. 7–10. https://doi.org/10.1109/ICOT.2017.8336075
49. "Child maltreatment." (n.d.) https://www.who.int/en/news-room/fact-sheets/detail/child-maltreatment. Accessed 09 Nov 2021
50. Chua CS, Guan H, Ho YK (2002) Model-based 3D hand posture estimation from a single 2D image. Image Vis Comput 20(3):191–202. https://doi.org/10.1016/S0262-8856(01)00094-4
51. Chung YY (2021) Design and Implementation of CNN-Based Human Activity Recognition System using WiFi Signals. J Adv Navig Technol 25(4):299–304. https://doi.org/10.12673/JANT.2021.25.4.299
52. "Combining CNN streams of RGB-D and skeletal data for human activity recognition | Elsevier Enhanced Reader." (n.d.) https://reader.elsevier.com/reader/sd/pii/S0167865518301636?token=783B2B6816D52EBDE82954EF671CD3D613E82F9615D667D214FBAE255C11798716F7DDC0FDCCC7E62D77F1BE4C2CBBEC&originRegion=eu-west-1&originCreation=20210925093934. Accessed 25 Sep 2021
53. Dai R et al (2022) Toyota smarthome untrimmed: real-world untrimmed videos for activity detection. IEEE Trans Pattern Anal Mach Intell 45:2533–2550. https://doi.org/10.1109/TPAMI.2022.3169976
54. Damaševičius R, Vasiljevas M, Šalkevičius J, Woźniak M (2016) Human activity recognition in AAL environments using random projections. Comput Math Methods Med 2016:1–17. https://doi.org/10.1155/2016/4073584
55. Dang X, Huang Y, Hao Z, Si X (2018) PCA-Kalman: device-free indoor human behavior detection with commodity Wi-Fi. EURASIP J Wirel Commun Netw 2018(1):1–17. https://doi.org/10.1186/S13638-018-1230-2/FIGURES/15
56. Das Dawn D, Shaikh SH (2015) A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector. Vis Comput 32(3):289–306. https://doi.org/10.1007/S00371-015-1066-2
57. Dash SCB, Mishra SR, Srujan Raju K, Narasimha Prasad LV (2021) Human action recognition using a hybrid deep learning heuristic. Soft Comput 25(20):13079–13092. https://doi.org/10.1007/S00500-021-06149-7
58. Wang J, Liu Z, Wu Y, Yuan J (2012) Mining action let ensemble for action recognition with depth cameras, IEEE Conf Comput Vis Pattern Recognit (CVPR 2012), Providence, Rhode Island, June 16–21
59. Dhiman C, Vishwakarma DK (2020) View-Invariant Deep Architecture for Human Action Recognition Using Two-Stream Motion and Shape Temporal Dynamics. IEEE Trans Image Process 29:3835–3844. https://doi.org/10.1109/TIP.2020.2965299
60. Dollár P, Rabaud V, Cottrell G, Belongie S (2005) Behavior recognition via sparse spatio-temporal features. *Proceedings - 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, VS-PETS*, vol. 2005, pp. 65–72. https://doi.org/10.1109/VSPETS.2005.1570899
61. Du Y, Chen F, Xu W (2007) Human interaction representation and recognition through motion decomposition. IEEE Signal Process Lett 14(12):952–955. https://doi.org/10.1109/LSP.2007.908035
62. Duan H, Lin K, Jin S, Liu W, Qian C, Ouyang W TRB: A Novel Triplet Representation for Understanding 2D Human Body
63. Duong TV, Bui HH, Phung DQ, Venkatesh S (2005) Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model
64. Duong TV, Phung DQ, Bui HH, Venkatesh S (n.d.) Human Behavior Recognition with Generic Exponential Family Duration Modeling in the Hidden Semi-Markov Model
65. Edwards M, Deng J, Xie X (2016) From pose to activity: Surveying datasets and introducing CONVERSE. Comput Vis Image Underst 144:73–105. https://doi.org/10.1016/J.CVIU.2015.10.010
66. El S, el Moudden I, Rabat I, Ouzir MM, Benyacoub B, el Bernoussi S (2016) Mining Human Activity Using Dimensionality Reduction and Pattern Recognition. Contemp Eng Sci 9(21):1031–1041. https://doi.org/10.12988/ces.2016.67119
67. Espinosa R, Ponce H, Gutiérrez S, Martínez-Villaseñor L, Brieva J, Moya-Albor E (2019) A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case

study using the UP-Fall detection dataset. Comput Biol Med 115:103520. https://doi.org/10.1016/J.COMPBIOMED.2019.103520
68. Fathi A, Ren X, Rehg JM (2011) Learning to recognize objects in egocentric activities. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3281–3288. https://doi.org/10.1109/CVPR.2011.5995444
69. Feng W et al (2014) Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera. SIViP 8:1129–1138. https://doi.org/10.1007/s11760-014-0645-4
70. "FineGym: A Hierarchical Video Dataset for Fine-grained Action Understanding." (n.d.) https://sdolivia.github.io/FineGym/ Accessed 26 Aug 2021
71. Foroughi H, Rezvanian A, Paziraee A (2008) Robust fall detection using human shape and multi-class support vector machine. *Proceedings - 6th Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP 2008*, pp. 413–420, https://doi.org/10.1109/ICVGIP.2008.49
72. Foroughi H, Aski BS, Pourreza H (2008) Intelligent video surveillance for monitoring fall detection of elderly in home environments. *Proceedings of 11th International Conference on Computer and Information Technology, ICCIT 2008*, pp. 219–224. https://doi.org/10.1109/ICCITECHN.2008.4803020
73. Foroughi H, Yazdi HS, Pourreza H, Javidi M (2008) An eigenspace-based approach for human fall detection using integrated time motion image and multi-class support vector machine. Proceedings - 2008 IEEE 4th International Conference on Intelligent Computer Communication and Processing, ICCP 2008. 83–90. https://doi.org/10.1109/ICCP.2008.4648358
74. Foroughi H, Naseri A, Saberi A, Sadoghi Yazdi H (2008) An eigenspace-based approach for human fall detection using integrated time motion image and neural network. 2008 9th international conference on signal processing, Beijing, China, 2008, pp. 1499–1503. https://doi.org/10.1109/ICOSP.2008.4697417
75. Franco A, Magnani A, Maio D (2020) A multimodal approach for human activity recognition based on skeleton and RGB data. Pattern Recogn Lett 131:293–299. https://doi.org/10.1016/J.PATREC.2020.01.010
76. Fu B, Kirchbuchner F, Kuijper A, Braun A, Gangatharan DV (2018) Fitness Activity Recognition on Smartphones Using Doppler Measurements. Informatics 5:24. https://doi.org/10.3390/INFORMATICS5020024
77. Gao J, Hauptmann AG, Bharucha A, Wactlar HD (2004) Dining activity analysis using a hidden Markov model. Proceedings of the 17th international conference on pattern recognition 2004, Cambridge, UK, 2:915–918. https://doi.org/10.1109/ICPR.2004.1334408
78. Ghali A, Cunningham AS, Pridmore TP (2003) Object and event recognition for stroke rehabilitation. Vis Commun Image Process 5150:980–989. https://doi.org/10.1117/12.503470
79. Ghazal S, Khan US, Saleem MM, Rashid N, Iqbal J (2019) Human activity recognition using 2D skeleton data and supervised machine learning. IET Image Process 13(13):2572–2578. https://doi.org/10.1049/IET-IPR.2019.0030
80. Ghorbani S et al (2021) MoVi: A large multi-purpose human motion and video dataset. PLoS One 16(6). https://doi.org/10.1371/JOURNAL.PONE.0253157
81. Goffredo M, Schmid M, Conforto S, Carli M, Neri A, D'Alessio T (2009) Markerless human motion analysis in Gauss-Laguerre transform domain: An application to sit-to-stand in young and elderly people. IEEE Trans Inf Technol Biomed 13(2):207–216. https://doi.org/10.1109/TITB.2008.2007960
82. "Google AI Blog: On-device, Real-time Body Pose Tracking with MediaPipe BlazePose." (n.d.) https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html. Accessed 14 Dec 2021
83. Gu Y, Ren F, Li J (2016) PAWS: Passive Human Activity Recognition Based on WiFi Ambient Signals. IEEE Internet Things J 3(5):796–805. https://doi.org/10.1109/JIOT.2015.2511805
84. Gu F, Chung MH, Chignell M, Valaee S, Zhou B, Liu X (2021) A Survey on Deep Learning for Human Activity Recognition. ACM Computing Surveys (CSUR) 54(8). https://doi.org/10.1145/3472290
85. Guo G, Lai A (2014) A survey on still image based human action recognition. Pattern Recogn 47(10):3343–3361. https://doi.org/10.1016/J.PATCOG.2014.04.018
86. Gupta JP, Singh N, Dixit P, Semwal VB, Dubey SR (2013) Human Activity Recognition Using Gait Pattern. Int J Comput Vis Image Process 3(3):31–53. https://doi.org/10.4018/IJCVIP.2013070103
87. Harrou F, Zerrouki N, Sun Y, Houacine A (2017) Vision-based fall detection system for improving safety of elderly people. IEEE Instrum Meas Mag 20(6):49–55. https://doi.org/10.1109/MIM.2017.8121952
88. "Home Action Genome." (n.d.) https://homeactiongenome.org/. Accessed 29 Jan 2023
89. Hu G, Huang S, Zhao L, Alempijevic A, Dissanayake G (2012) A robust RGB-D SLAM algorithm. *IEEE International Conference on Intelligent Robots and Systems*, pp. 1714–1719. https://doi.org/10.1109/IROS.2012.6386103

90. Xu F, Fujimura K (2003) Human detection using depth and gray images. Proceedings - IEEE conference on advanced video and signal based surveillance, AVSS 2003:115–121. https://doi.org/10.1109/AVSS.2003.1217910
91. Huo F, Hendriks E, Paclik P, Oomes AHJ (2009) Markerless human motion capture and pose recognition. *2009 10th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2009*, pp. 13–16. https://doi.org/10.1109/WIAMIS.2009.5031420
92. Ihianle IK, Nwajana AO, Ebenuwa SH, Otuka RI, Owa K, Orisatoki MO (2020) A deep learning approach for human activities recognition from multimodal sensing devices. IEEE Access 8:179028–179038. https://doi.org/10.1109/ACCESS.2020.3027979
93. "II. DESCRIPTION OF WORK". (n.d.)
94. Schreier H, Orteu JJ, Sutton MA (2009) Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications. Image correlation for shape, motion and deformation measurements: basic concepts,theory and applications, pp. 1–321. https://doi.org/10.1007/978-0-387-78747-3/COVER
95. Incel OD, Ozgovde A (2018) ARService: A Smartphone based Crowd-Sourced Data Collection and Activity Recognition Framework. Procedia Comput Sci 130:1019–1024. https://doi.org/10.1016/J.PROCS.2018.04.142
96. "Introducing the Penn Action Dataset | Penn Action." (n.d.) http://dreamdragon.github.io/PennAction/. Accessed 26 Aug 2021
97. Islam MM, Iqbal T (Oct. 2020) HAMLET: A hierarchical multimodal attention-based human activity recognition algorithm. *IEEE International Conference on Intelligent Robots and Systems*, pp. 10285–10292. https://doi.org/10.1109/IROS45743.2020.9340987
98. Jalal A, Kim Y (Oct. 2014) Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data. *11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014*, pp. 119–124. https://doi.org/10.1109/AVSS.2014.6918654
99. Jalal A, Uddin I (2007) Security architecture for third generation (3G) using GMHS cellular network. *Proceedings - 3rd International Conference on Emerging Technologies, ICET 2007*, pp. 74–79. https://doi.org/10.1109/ICET.2007.4516319.
100. Jalal A, Lee S, Kim JT, Kim T-S (2012) Human Activity Recognition via the Features of Labeled Depth Body Parts. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7251 LNCS, pp. 246–249. https://doi.org/10.1007/978-3-642-30779-9_36
101. Jalal A, Kamal S, Kim D (2014) A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments. Sensors 14(7):11735–11759. https://doi.org/10.3390/S140711735
102. Jalal A, Kamal S, Kim D (2016) Human Depth Sensors-Based Activity Recognition Using Spatiotemporal Features and Hidden Markov Model for Smart Environments. J Comput Netw Commun 2016:1–11. https://doi.org/10.1155/2016/8087545
103. Jalal A, Kim YH, Kim YJ, Kamal S, Kim D (2017) Robust human activity recognition from depth video using spatiotemporal multi-fused features. Pattern Recogn 61:295–308. https://doi.org/10.1016/J.PATCOG.2016.08.003
104. Jaouedi N, Perales FJ, Buades JM, Boujnah N, Bouhlel MS (2020) Prediction of Human Activities Based on a New Structure of Skeleton Features and Deep Learning Model. Sensors 20:4944. https://doi.org/10.3390/s20174944
105. Kang S-M, Wildes R (2016) Review of Action Recognition and Detection Methods. *undefined*
106. Kareem I, Ali SF, Sheharyar A (Nov. 2020) Using Skeleton based Optimized Residual Neural Network Architecture of Deep Learning for Human Fall Detection. *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*. https://doi.org/10.1109/INMIC50486.2020.9318061
107. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Li FF (Sep. 2014) Large-scale video classification with convolutional neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732. https://doi.org/10.1109/CVPR.2014.223
108. Ke Y, Sukthankar R, Hebert M (2007) Spatio-temporal shape and flow correlation for action recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, https://doi.org/10.1109/CVPR.2007.383512
109. Khan ZA, Sohn W (2011) Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care. IEEE Trans Consum Electron 57(4):1843–1850. https://doi.org/10.1109/TCE.2011.6131162
110. "Kinetics | DeepMind." (n.d.) https://deepmind.com/research/open-source/kinetics. Accessed 26 Aug 2021

111. Kotsiantis S, Kanellopoulos D, Pintelas P (2007) Data Preprocessing for Supervised Leaning. *undefined*
112. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst 25 [Online]. Available: http://code.google.com/p/cuda-convnet/. Accessed 06 Oct 2021
113. Kumari S, Mitra SK (2011) Human action recognition using DFT. *Proceedings - 2011 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2011*, pp. 239–242, https://doi.org/10.1109/NCVPRIPG.2011.58
114. Kuo Y-M, Lee J-S, Chung P-C (2010) A visual context-awareness-based sleeping-respiration measurement system. IEEE Trans Inf Technol Biomed 14(2):255–265. https://doi.org/10.1109/TITB.2009.2036168
115. Lahiri D, Dhiman C, Vishwakarma DK (Apr. 2018) Abnormal human action recognition using average energy images. *2017 Conference on Information and Communication Technology, CICT 2017*, vol. 2018-April, pp. 1–5. https://doi.org/10.1109/INFOCOMTECH.2017.8340622
116. Laptev I, Marszałek M, Schmid C, Rozenfeld B (2008) Learning realistic human actions from movies. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587756
117. Lee TK, Lim S, Lee S, An S, Oh SY (2012) Indoor mapping using planes extracted from noisy RGB-D sensors. *IEEE International Conference on Intelligent Robots and Systems*, pp. 1727–1733. https://doi.org/10.1109/IROS.2012.6385909
118. Leu A, Ristic-Durrant D, Graser A (2011) A robust markerless vision-based human gait analysis system. *SACI 2011 - 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*, pp. 415–420. https://doi.org/10.1109/SACI.2011.5873039
119. Li W, Zhang Z, Liu Z (2010) Action recognition based on a bag of 3D points. IEEE computer society conference on computer vision and pattern recognition - workshops, CVPRW 2010, pp. 9–14. https://doi.org/10.1109/CVPRW.2010.5543273
120. Li YR, Miaou SG, Hung CK, Sese JT (2011) A gait analysis system using two cameras with orthogonal view. *2011 International Conference on Multimedia Technology, ICMT 2011*, pp. 2841–2844. https://doi.org/10.1109/ICMT.2011.6002046
121. Li C, Tong R, Tang M (2018) Modelling Human Body Pose for Action Recognition Using Deep Neural Networks. Arab J Sci Eng 43(12):7777–7788. https://doi.org/10.1007/S13369-018-3189-Z
122. Li X, Dong W, Shi G (2018) Sparsity-Based Denoising of Photographic Images: From Model-Based to Data-Driven. Adv Comput Vis Pattern Recognit:37–62. https://doi.org/10.1007/978-3-319-96029-6_2
123. Liangliang W, Lianzheng G, Ruifeng L, Yajun F (2017) Three-stream CNNs for action recognition. Pattern Recogn Lett 92:33–40. https://doi.org/10.1016/J.PATREC.2017.04.004
124. Liao TY, Miaou SG, Li YR (2010) A vision-based walking posture analysis system without markers. *ICSPS 2010 - Proceedings of the 2010 2nd International Conference on Signal Processing Systems*, vol. 3. https://doi.org/10.1109/ICSPS.2010.5555656
125. Lin C-H, Hsu F-S, Lin W-Y (2010) Recognizing Human Actions Using NWFE-Based Histogram Vectors. EURASIP J Adv Signal Process 2010(1):1–15. https://doi.org/10.1155/2010/453064
126. Lin L, Wang K, Zuo W, Wang M, Luo J, Zhang L (Dec. 2015) A Deep Structured Model with Radius-Margin Bound for 3D Human Activity Recognition. [Online]. Available: http://arxiv.org/abs/1512.01642. Accessed 28 Sep 2021
127. Lin W, Hasenstab K, Moura Cunha G, Schwartzman A (2020) Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. Sci Rep 10(1):1–11. https://doi.org/10.1038/s41598-020-77264-y
128. Liu L, Shao L (n.d.) Learning Discriminative Representations from RGB-D Video Data
129. Liu C-D, Chung P, Chung Y, Thonnat M (2007) Understanding of human behaviors from videos in nursing care monitoring systems. *undefined*
130. Liu C, Hu Y, Li Y, Song S, Liu J (Mar. 2017) PKU-MMD: A Large Scale Benchmark for Continuous Multi-Modal Human Action Understanding. [Online]. Available: http://arxiv.org/abs/1703.07475. Accessed 23 Nov 2021
131. Lühr S, Venkatesh S, West G, Bui HH (n.d.) Explicit State Duration HMM for Abnormality Detection in Sequences of Human Activity
132. Lun R, Zhao W (2015) A survey of applications and human motion recognition with Microsoft Kinect. Int J Pattern Recognit Artif Intell 29(5). https://doi.org/10.1142/S0218001415550083
133. Luo Y, der Wu T, Hwang JN (2003) Object-based analysis and interpretation of human motion in sports video sequences by dynamic bayesian networks. Comput Vis Image Underst 92(2–3):196–216. https://doi.org/10.1016/J.CVIU.2003.08.001

134. Lygouras E, Santavas N, Taitzoglou A, Tarchanidis K, Mitropoulos A, Gasteratos A (2019) Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations. Sensors 19(16):3542. https://doi.org/10.3390/S19163542
135. Manjarres J, Lan G, Gorlatova M, Hassan M, Pardo M (2022) Deep learning for detecting human activities from piezoelectric-based kinetic energy signals. IEEE Internet Things J 9(10):7545–7558. https://doi.org/10.1109/JIOT.2021.3093245
136. Marinho LB, de Souza Junior AH, Filho PPR (2016) A New Approach to Human Activity Recognition Using Machine Learning Techniques. Adv Intell Syst Comput 557:529–538. https://doi.org/10.1007/978-3-319-53480-0_52
137. Mekruksavanich S, Promsakon C, Jitpattanakul A (2021) Location-based daily human activity recognition using hybrid deep learning network. JCSSE 2021 - 18th international joint conference on computer science and software engineering: cybernetics for human beings. https://doi.org/10.1109/JCSSE53117.2021.9493807
138. Meng B, Liu X, Wang X (2018) Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos. Multimed Tools Appl 77(20):26901–26918. https://doi.org/10.1007/S11042-018-5893-9
139. Menicatti R, Bruno B, Sgorbissa A (2017) Modelling the influence of cultural information on vision-based human home activity recognition. 2017 14th international conference on ubiquitous robots and ambient intelligence, URAI 2017, pp. 32–38. https://doi.org/10.1109/URAI.2017.7992880
140. Seidenari L, Varano V, Berretti S, Del Bimbo A, Pala P (2013) Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses", 3rd International Workshop on Human Activity Understanding from 3D data (HAU3D'13), in conjunction with CVPR 2013, Portland, Oregon
141. Minh Dang L, Min K, Wang H, Jalil Piran M, Hee Lee C, Moon H (2020) Sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Recogn 108:107561. https://doi.org/10.1016/J.PATCOG.2020.107561
142. Mohammed Hashim BA, Amutha R (2020) Human activity recognition based on smartphone using fast feature dimensionality reduction technique. J Ambient Intell Humaniz Comput (JAIHC) 12(2):2365–2374. https://doi.org/10.1007/S12652-020-02351-X
143. Zhang L, Halber M, Rusinkiewicz S (2019) Accelerating large-kernel convolution using summed-area tables. arXiv preprint arXiv:1906.11367
144. Mukherjee S, Anvitha L, Lahari TM (2020) Human activity recognition in RGB-D videos by dynamic images. Multimed Tools Appl 79(27):19787–19801. https://doi.org/10.1007/S11042-020-08747-3
145. Natarajan P, Nevatia R (2008) Online, real-time tracking and recognition of human actions (2008) IEEE workshop on motion and video computing. WMVC 2008. https://doi.org/10.1109/WMVC.2008.4544064
146. Ni T, Chen Y, Song K, Xu W (2021) A simple and fast human activity recognition system using radio frequency energy harvesting. UbiComp/ISWC 2021 - adjunct proceedings of the 2021 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2021. ACM International Symposium on Wearable Computers, pp. 666–671. https://doi.org/10.1145/3460418.3480399
147. Niu W, Long J, Han D, Wang YF (2004) Human activity detection and recognition for video surveillance. *2004 IEEE International Conference on Multimedia and Expo (ICME)*, vol. 1, pp. 719–722. https://doi.org/10.1109/ICME.2004.1394293
148. Noori FM, Wallace B, Uddin MdZ, Torresen J (Jun. 2019) A Robust Human Activity Recognition Approach Using OpenPose, Motion Features, and Deep Recurrent Neural Network. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11482 LNCS, pp. 299–310. https://doi.org/10.1007/978-3-030-20205-7_25
149. Núñez-Marcos A, Azkune G, Arganda-Carreras I (2017) Vision-based fall detection with convolutional neural networks. Wirel Commun Mob Comput 2017. https://doi.org/10.1155/2017/9474806
150. Ofli F, Chaudhry R, Kurillo G, Vidal R, Bajcsy R (2013) Berkeley MHAD: a comprehensive multimodal human action database. Proceedings of IEEE workshop on applications of computer vision, pp. 53–60. https://doi.org/10.1109/WACV.2013.6474999
151. Oikonomidis I, Kyriazis N, Argyros A (2011) Efficient model-based 3D tracking of hand articulations using Kinect pp. 101.1–101.11. https://doi.org/10.5244/C.25.101
152. Okeyo G, Chen L, Wang H (2013) An agent-mediated ontology-based approach for composite activity recognition in smart homes. J Univ Comput Sci 19(17):2577–2597. https://doi.org/10.3217/JUCS-019-17-2577
153. Paoletti G, Cavazza J, Beyan C, del Bue A (Apr. 2022) Unsupervised Human Action Recognition with Skeletal Graph Laplacian and Self-Supervised Viewpoints Invariance. https://doi.org/10.48550/arxiv.2204.10312

154. Perez-Sala X, Escalera S, Angulo C (2012) Survey on 2D and 3D human pose recovery. Front Artif Intell Appl 248:101–110. https://doi.org/10.3233/978-1-61499-139-7-101
155. Perry S (2018) Image and Video Noise: An Industry Perspective. Adv Comput Vis Pattern Recognit:207–234. https://doi.org/10.1007/978-3-319-96029-6_8
156. Petscharnig S, Lux M, Chatzichristofis S (Jun. 2017) Dimensionality reduction for image features using deep learning and autoencoders. *ACM International Conference Proceeding Series*, vol. Part F130150. https://doi.org/10.1145/3095713.3095737
157. Pham HH, Salmane H, Khoudour L, Crouzil A, Zegers P, Velastin SA (2019) A Unified Deep Framework for Joint 3D Pose Estimation and Action Recognition from a Single RGB Camera. Sensors (Switzerland) 20(7) [Online]. Available: https://arxiv.org/abs/1907.06968v1. Accessed 30 Oct 2021
158. "Prepare the 20BN-something-something Dataset V2 — gluoncv 0.11.0 documentation." (n.d.) https://cv.gluon.ai/build/examples_datasets/somethingsomethingv2.html. Accessed 26 Aug 2021
159. Rahman M, Das T (2021) Human activity recognition using deep learning-based approach. Lecture Notes in Networks and Systems 204:813–830. https://doi.org/10.1007/978-981-16-1089-9_63/COVER
160. Ramanathan M, Yau WY, Teoh EK (2014) Human action recognition with video data: Research and evaluation challenges. IEEE Trans Hum Mach Syst 44(5):650–663. https://doi.org/10.1109/THMS.2014.2325871
161. Ray S, Alshouiliy K, Agrawal DP (2020) Dimensionality Reduction for Human Activity Recognition Using Google Colab. Information 12(1):6. https://doi.org/10.3390/INFO12010006
162. Kang SM, Wildes RP (2016) Review of action recognition and detection methods. arXiv preprint arXiv:1610.06906
163. "Recognizing Non-rigid Human Actions Using Joints Tracking in Space-time | Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) Volume 2 - Volume 2." (n.d.) https://dl.acm.org/doi/abs/10.5555/977403.978309. Accessed 19 Aug 2021
164. Sani S, Wiratunga N, Massie S, Cooper K (2017) kNN Sampling for Personalised Human Activity Recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10339 LNAI, pp. 330–344. https://doi.org/10.1007/978-3-319-61030-6_23
165. Schrader L et al (2020) Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people. J Popul Ageing 13(2):139–165. https://doi.org/10.1007/S12062-020-09260-Z
166. Schüldt C, Laptev I, Caputo B (2004) Recognizing human actions: A local SVM approach. Proc - Int Conf Pattern Recog 3:32–36. https://doi.org/10.1109/ICPR.2004.1334462
167. Scovanner P, Ali S, Shah M (2007) A 3-dimensional sift descriptor and its application to action recognition. Proceedings of the ACM International Multimedia Conference and Exhibition, pp. 357–360. https://doi.org/10.1145/1291233.1291311
168. Sempena S, Maulidevi NU, Aryan PR (2011) Human action recognition using Dynamic Time Warping. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011*, https://doi.org/10.1109/ICEEI.2011.6021605
169. Wang L, Qiao Y, Tang X (2015) Action recognition with trajectory-pooled deep-convolutional descriptors. In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 4305–4314
170. Seshan A (n.d.) Enabling High-Accuracy Human Activity Recognition with Fine-Grained Indoor Localization
171. Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. Adv Neural Inf Process Syst vol. 1, pp. 568–576. Available: https://arxiv.org/abs/1406.2199v2. Accessed 7 May 2023
172. Singh T, Vishwakarma DK (2021) A deep multimodal network based on bottleneck layer features fusion for action recognition. Multimed Tools Appl 2021:1–21. https://doi.org/10.1007/S11042-021-11415-9
173. Soomro K, Zamir AR, Shah M (Dec. 2012) UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. [Online]. Available: https://arxiv.org/abs/1212.0402v1. Accessed 26 Aug 2021
174. Yoshikawa Y, Lin J, Takeuchi A (2018) Stair actions: a video dataset of everyday home actions. arXiv preprint arXiv:1804.04326.
175. Subetha T, Chitrakala S (Jul. 2016) A survey on human activity recognition from videos. *2016 International Conference on Information Communication and Embedded Systems, ICICES 2016*, https://doi.org/10.1109/ICICES.2016.7518920
176. Sumiya T, Matsubara Y, Nakano M, Sugaya M (2015) A Mobile Robot for Fall Detection for Elderly-Care. Procedia Comput Sci 60(1):870–880. https://doi.org/10.1016/J.PROCS.2015.08.250

177. Sun Z, Ke Q, Rahmani H, Bennamoun M, Wang G, Liu J (2022) Human Action Recognition From Various Data Modalities: A Review. IEEE Trans Pattern Anal Mach Intell 45:1–20. https://doi.org/10.1109/TPAMI.2022.3183112
178. Sung J, Ponce C, Selman B, Saxena A (Jul. 2011) Unstructured Human Activity Detection from RGBD Images. *Proc IEEE Int Conf Robot Autom*, pp. 842–849. https://doi.org/10.1109/icra.2012.6224591
179. Tao D, Jin L, Yang Z, Li X (2013) Rank preserving sparse learning for kinect based scene classification. IEEE Trans Cybern 43(5):1406–1417. https://doi.org/10.1109/TCYB.2013.2264285
180. Tao D, Cheng J, Lin X, Yu J (2015) Local structure preserving discriminative projections for RGB-D sensor-based scene classification. Inf Sci (N Y) 320:383–394. https://doi.org/10.1016/J.INS.2015.03.031
181. "Three Major Challenges Facing IoT - IEEE Internet of Things." (n.d.) https://iot.ieee.org/newsletter/march-2017/three-major-challenges-facing-iot.html/. Accessed 20 Dec 2021
182. Tiwari G, Bajaj P, Gupta S (2021) mmFiT: Contactless Fitness Tracker Using mmWave Radar and Edge Computing Enabled Deep Learning. IEEE Internet Things J (Y). https://doi.org/10.36227/TECHRXIV.16574588.V1
183. Uddin MA, Lee Y-K (2019) Feature Fusion of Deep Spatial Features and Handcrafted Spatiotemporal Features for Human Action Recognition. Sensors 19(7):1599. https://doi.org/10.3390/S19071599
184. "UTKinect-Action3D Dataset." (n.d.) http://cvrc.ece.utexas.edu/KinectDatasets/HOJ3D.html. Accessed 23 Nov 2021
185. Vallathan G, John A, Thirumalai C, Mohan S, Srivastava G, Lin JC-W (2020) Suspicious activity detection using deep learning in secure assisted living IoT environments. J Supercomput 77(4):3242–3260. https://doi.org/10.1007/S11227-020-03387-8
186. Vandersmissen B, Knudde N, Jalalvand A, Couckuyt I, Dhaene T, de Neve W (2020) Indoor human activity recognition using high-dimensional sensors and deep neural networks. Neural Comput Appl 32(16):12295–12309. https://doi.org/10.1007/S00521-019-04408-1/TABLES/8
187. Veeraraghavan A, Roy-Chowdhury AK, Chellappa R (2005) Matching shape sequences in video with applications in human movement analysis. IEEE Trans Pattern Anal Mach Intell 27(12):1896–1909. https://doi.org/10.1109/TPAMI.2005.246
188. Vishwakarma S, Agrawal A (2012) A survey on activity recognition and behavior understanding in video surveillance. Vis Comput 29(10):983–1009. https://doi.org/10.1007/S00371-012-0752-6
189. Vishwakarma S, Agrawal A (2013) A survey on activity recognition and behavior understanding in video surveillance. Vis Comput 29(10):983–1009. https://doi.org/10.1007/S00371-012-0752-6/EMAIL/CORRESPONDENT/C1/NEW
190. Vrigkas M, Nikou C, Kakadiaris IA (2015) A Review of Human Activity Recognition Methods. Front Robot AI 2(NOV):28. https://doi.org/10.3389/FROBT.2015.00028
191. Wan S, Qi L, Xu X, Tong C, Gu Z (2019) Deep Learning Models for Real-time Human Activity Recognition with Smartphones. Mob Netw Appl 25(2):743–755. https://doi.org/10.1007/S11036-019-01445-X
192. Wang J, Liu Z, Wu Y, Yuan J (2012) Mining actionlet ensemble for action recognition with depth cameras. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1290–1297. https://doi.org/10.1109/CVPR.2012.6247813
193. Wang W, Huang Y, Wang Y, Wang L (2014) Generalized Autoencoder: A Neural Network Framework for Dimensionality Reduction. pp. 490–497.
194. Wang L, Qiao Y, Tang X (May 2015) Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07–12-June-2015, pp. 4305–4314. https://doi.org/10.1109/cvpr.2015.7299059
195. Wang Y, Yao H, Zhao S (2016) Auto-encoder based dimensionality reduction. Neurocomputing 184:232–242. https://doi.org/10.1016/J.NEUCOM.2015.08.104
196. Wang P, Cao Y, Shen C, Liu L, Shen HT (n.d.) TEMPORAL PYRAMID POOLING CNN FOR ACTION RECOGNITION 1 Temporal Pyramid Pooling Based Convolutional Neural Network for Action Recognition
197. "Wearable Health Device Dermatitis: A Case of Acrylate-Related Contact Allergy | MDedge Dermatology." (n.d.) https://www.mdedge.com/dermatology/article/143798/contact-dermatitis/wearable-health-device-dermatitis-case-acrylate. Accessed 30 Oct 2021
198. WHO global report on falls prevention in older age." https://www.who.int/publications/i/item/9789241563536. Accessed 7 May 2023

199. Wu C, Zhang J, Sener O, Selman B, Savarese S, Saxena A (2016) Watch-n-Patch: Unsupervised Learning of Actions and Relations. IEEE Trans Pattern Anal Mach Intell 40(2):467–481 [Online]. Available: https://arxiv.org/abs/1603.03541v1. Accessed 26 Aug 2021
200. Xiang D, Joo H, Sheikh Y (Dec. 2018) Monocular Total Capture: Posing Face, Body, and Hands in the Wild. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 10957–10966. [Online]. Available: https://arxiv.org/abs/1812.01598v1. Accessed 25 Sep 2021
201. Xiao Y, Xing C, Zhang T, Zhao Z (2019) An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks. IEEE Access 7:42210–42219. https://doi.org/10.1109/ACCESS.2019.2904620
202. Yadav SK, Tiwari K, Pandey HM, Akbar SA (2021) A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowl Based Syst 223:106970. https://doi.org/10.1016/J.KNOSYS.2021.106970
203. "Yamato-HumanAction". (n.d.)
204. Yang X, Tian Y (2014) Super Normal Vector for Activity Recognition Using Depth Sequences. pp. 804–811
205. Yang W, Liu X, Zhang L, Yang LT (2013) Big data real-time processing based on storm. *Proceedings - 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2013*, pp. 1784–1787. https://doi.org/10.1109/TRUSTCOM.2013.247
206. Yang L, Ren Y, Zhang W (2016) 3D depth image analysis for indoor fall detection of elderly people. Digital Commun Netw 2(1):24–34. https://doi.org/10.1016/J.DCAN.2015.12.001
207. Yu M, Gong L, Kollias S (Nov. 2017) Computer vision based fall detection by a convolutional neural network. *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, vol. 2017-January, pp. 416–420. https://doi.org/10.1145/3136755.3136802
208. Zanfir M, Leordeanu M, Sminchisescu C (2013) The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection. https://doi.org/10.1109/ICCV.2013.342
209. Zehra N, Azeem SH, Farhan M (Mar. 2021) Human activity recognition through ensemble learning of multiple convolutional neural networks. *2021 55th Annual Conference on Information Sciences and Systems, CISS 2021*. https://doi.org/10.1109/CISS50987.2021.9400290
210. Zhan K, Faux S, Ramos F (2015) Multi-scale Conditional Random Fields for first-person activity recognition on elders and disabled patients. Pervasive Mob Comput 16(PB):251–267. https://doi.org/10.1016/j.pmcj.2014.11.004
211. Zhang C, Tian Y, Capezuti E (2012) Privacy Preserving Automatic Fall Detection for Elderly Using RGBD Cameras. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7382 LNCS, no. PART 1, pp. 625–633. https://doi.org/10.1007/978-3-642-31522-0_95
212. Zhang Z, Conly C, Athitsos V (Jul. 2015) A survey on vision-based fall detection. *8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2015 – Proceedings*. https://doi.org/10.1145/2769493.2769540
213. Zhao H, Torralba A, Torresani L, Yan Z (Oct. 2019) HACS: Human action clips and segments dataset for recognition and temporal localization. *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 8667–8677
214. Zhu S (2021) Multiple Target Tracking and Human Activity Recognition based on The IR-UWB Radar Sensor Networks. [Online]. Available: https://repository.tudelft.nl/islandora/object/uuid%3A8a0a66cc-7d94-4149-aa56-4213a588b86d. Accessed 17 Oct 2021
215. Zhu D, Pang N, Li G, Liu S (Jun. 2017) NotiFi: A ubiquitous WiFi-based abnormal activity detection system. *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, pp. 1766–1773. https://doi.org/10.1109/IJCNN.2017.7966064
216. Ziaeefard M, Bergevin R (2015) Semantic human activity recognition: A literature review. Pattern Recogn 48(8):2329–2345. https://doi.org/10.1016/J.PATCOG.2015.03.006