



# Human activity recognition in artificial intelligence framework: a narrative review

Neha Gupta<sup>1,3</sup> · Suneet K. Gupta<sup>1</sup> · Rajesh K. Pathak<sup>2</sup> · Vanita Jain<sup>3</sup> · Parisa Rashidi<sup>4</sup> · Jasjit S. Suri<sup>5,6</sup> 

Published online: 18 January 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2021

## Abstract

Human activity recognition (HAR) has multifaceted applications due to its worldly usage of acquisition devices such as smartphones, video cameras, and its ability to capture human activity data. While electronic devices and their applications are steadily growing, the advances in Artificial intelligence (AI) have revolutionized the ability to extract deep hidden information for accurate detection and its interpretation. This yields a better understanding of rapidly growing acquisition devices, AI, and applications, the three pillars of HAR under one roof. There are many review articles published on the general characteristics of HAR, a few have compared all the HAR devices at the same time, and few have explored the impact of evolving AI architecture. In our proposed review, a detailed narration on the three pillars of HAR is presented covering the period from 2011 to 2021. Further, the review presents the recommendations for an improved HAR design, its reliability, and stability. Five major findings were: (1) HAR constitutes three major pillars such as devices, AI and applications; (2) HAR has dominated the healthcare industry; (3) Hybrid AI models are in their infancy stage and needs considerable work for providing the stable and reliable design. Further, these trained models need solid prediction, high accuracy, generalization, and finally, meeting the objectives of the applications without bias; (4) little work was observed in abnormality detection during actions; and (5) almost no work has been done in forecasting actions. We conclude that: (a) HAR industry will evolve in terms of the three pillars of electronic devices, applications and the type of AI. (b) AI will provide a powerful impetus to the HAR industry in future.

**Keywords** Human activity recognition · Sensor-based · Vision-based · Radio frequency-based identification · Device-free · Imaging · Deep learning · Machine learning · And hybrid models

## Abbreviations

*CIT	Citations
AAL	Ambient assistive living
ADL	Activity of daily living

---

✉ Jasjit S. Suri  
jasjit.suri@atheropoint.com

Extended author information available on the last page of the article

AI	Artificial intelligence
AP	Average pooling
AUC	Area under curve
BN	Batch normalization
BS	Batch size
CEL	Cross entropy loss
CNN	Convolution neural network
CONV	Convolution
CV	Cross validation
DL	Deep learning
DO	Drop out
DTW	Dynamic time warping
EER	Equal error rate
FC	Fully connected
GAN	Generative adversarial n/w
GFLOP	Giga floating point operations/sec
GPU	Graphics processing unit
GRU	Gated recurrent unit
HAR	Human activity recognition
KL	Kullback Lieblar
LOSO	Leave one subject out
LR	Learning rate
LSTM	Long short-term memory
MAE	Mean absolute error
mAPs	Mean average precision
ML	Machine learning
MP	Max pooling
MSE	Mean square error
PCA	Principal component analysis
RFID	Radio frequency identification
RNN	Recurrent neural network
RSSI	Received signal strength indicator
Sens	Sensitivity
SGD	Stochastic gradient descent
SGD	Stochastic gradient descent
SM	Softmax
Spec	Specificity
SVM	Support vector machine
TL	Transfer learning
VAE	Variational autoencoders
P	Precision
R	Recall
TP	True positive
TN	True negative
FP	False positive
FN	False negative
LHR	Likelihood ratio

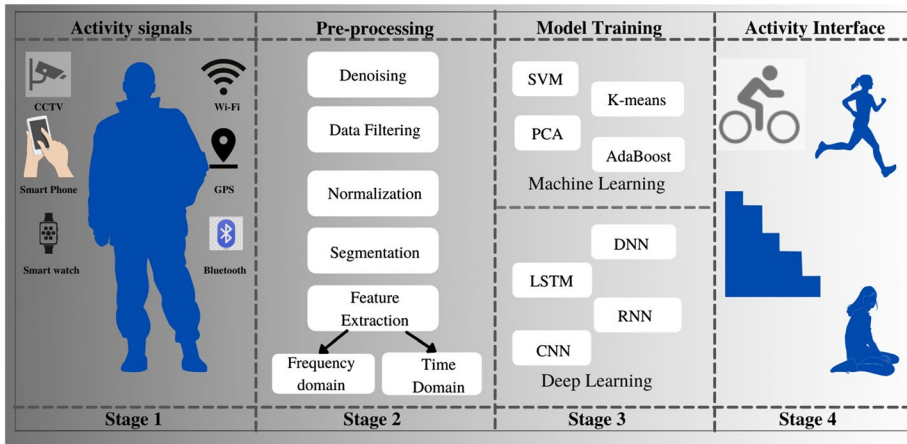
## 1 Introduction

Human activity recognition (HAR) can be referred to as the art of identifying and naming activities using Artificial Intelligence (AI) from the gathered activity raw data by utilizing various sources (so-called devices). Examples of such devices include wearable sensors (Pham et al. 2020), electronic device sensors like smartphone inertial sensor (Qi et al. 2018; Zhu et al. 2019), camera devices like Kinect (Wang et al. 2019a; Phyo et al. 2019), closed-circuit television (CCTV) (Du et al. 2019), and some commercial off-the-shelf (COTS) equipment's (Ding et al. 2015; Li et al. 2016). The use of diverse sources makes HAR important for multifaceted applications domains, such as healthcare (Pham et al. 2020; Zhu et al. 2019; Wang et al. 2018), surveillance (Thida et al. 2013; Deep and Zheng 2019; Vaniya and Bharathi 2016; Shuaibu et al. 2017; Beddiar et al. 2020) remote care to elderly people living alone (Phyo et al. 2019; Deep and Zheng 2019; Yao et al. 2018), smart home/office/city (Zhu et al. 2019; Deep and Zheng 2019; Fan et al. 2017), and various monitoring application like sports, and exercise (Ding et al. 2015). The widespread use of HAR is beneficial for the safety and quality of life for humans (Ding et al. 2015; Chen et al. 2020).

The existence of devices like sensors, video cameras, radio frequency identification (RFID), and Wi-Fi are not new, but the usage of these devices in HAR is in its infancy. The reason for HAR's evolution is the fast growth of techniques such as AI, which enables the use of these devices in various application domains (Suthar and Gadhia 2021). Therefore, we can say that there is a mutual relationship between the AI techniques or AI models and HAR devices. Earlier these models were based on a single image or a small sequence of images, but the advancements in AI have provided more opportunities. According to our observations (Chen et al. 2020; Suthar and Gadhia 2021; Ding et al. 2019), the growth of HAR is directly proportional to the advancement of AI which thrives the scope of HAR in various application domains.

The introduction of deep learning (DL) in the HAR domain has made the task of meaningful feature extraction from the raw sensor data. The evolution of DL models such as (1) convolutional neural networks (CNN) (Tandel et al. 2020), (2) extending the role of transfer weighting schemes (it allows the knowledge reusability where the recognition model is trained on a set of data and the same trained knowledge can then be used by a different testing dataset) such as Inception (Szegedy et al. 2015, 2016, 2017), VGG-16 (Simonyan and Zisserman 2015), and Residual Neural Network (Resnets)-50 (Nash et al. 2018), (3) series of hybrid DL models such as fusion of CNN with long short-term memory (LSTM), Inception with ResNets (Yao et al. 2017, 2019, 2018; Buffelli and Vandin 2020), (4) loss function designs such entropy, Kaulback Liberal divergence, and Tversky (Janocha and Czarnecki 2016; Wang et al. 2020a), (5) optimization paradigms such as cross-entropy, stochastic gradient descent (SGD) (Soydancer 2020; Sun et al. 2020) has made the task of HAR-based design plug-and-play based. Even though it is getting black-box oriented, it requires better understanding to actually ensure that the 3-legged stool is stable and effective.

Typically, HAR consists of four stages (Fig. 1) including (1) capturing of signal activity, (2) data pre-processing, (3) AI-based activity recognition, and (4) the user interface for the management of HAR. Each stage can be implemented using several techniques bringing the HAR system to have multiple choices. Thus, the choice of the application domain, the type of data acquisition device, and the processing of artificial intelligence (AI) algorithms for activity detection makes the choices even more challenging.



**Fig. 1** Four stages of HAR process (Hx et al. 2017)

Numerous reviews in HAR have been published, but our observations show that most of the studies are associated with either vision-based (Beddiar et al. 2020; Dhi-man Chhavi 2019; Ke et al. 2013) or sensor-based (Carvalho and Sofia 2020; Lima et al. 2019), while very few have considered RFID-based and device-free HAR. Further, there is no AI review article that covers the detailed analysis of all the four device types that includes all four types of devices such as sensor-based (Yao et al. 2017, 2019; Hx et al. 2017; Hsu et al. 2018; Xia et al. 2020; Murad and Pyun 2017), vision-based (Feichtenhofer et al. 2018; Simonyan and Zisserman 2014; Newell Alejandro 2016; Crasto et al. 2019), RFID-based (Han et al. 2014), and device-free (Zhang et al. 2011).

An important observation to note here is that technology has advanced in the field of AI, i.e., deep learning (Agarwal et al. 2021; Skandha et al. 2020; Saba et al. 2021) and machine learning methods (Hsu et al. 2018; Jamthikar et al. 2020) and is revolutionizing the ability to extract deep hidden information for accurate detection and interpretation. Thus, there is a need to understand the role of these new paradigms that are rapidly changing HAR devices. This puts the requirement to consider a review inclined to address simultaneously changing AI and HAR devices. Therefore, the main objective of this study is to better understand the HAR framework while integrating devices and application domains in the specialized AI framework. What types of devices can fit in which type of application, and what attributes of the AI can be considered during the design of such (Agarwal et al. 2021) a framework are some of the issues that need to be explored. Thus, this review is going to illustrate how one can select such a combination by first understanding the types of HAR devices, and then, the knowledge-based infrastructure in the fast-moving world of AI, knowing that some of such combinations can be transformed into different applications (domains).

The proposed review is structured as follows: Sect. 2 covers the search strategy, and literature review with statistical distributions of HAR attributes. Section 3 illustrates the description of the HAR stages, HAR devices, and HAR application domains in the AI framework. Section 4 illustrates the role of emerging AI as the core of HAR. Section 5 presents performance evaluation criteria in the HAR and integration of AI in HAR devices. Section 6 consists of a critical discussion on factors influencing HAR,

benchmarking of the study against the previous studies, and finally, the recommendations. Section 7 finally concludes the study.

## 2 Search strategy and literature review

“Google Scholar” is used for searching articles published between the periods of 2011-present. The search included the keywords “human activity recognition” or “HAR” in combination with terms “machine learning”, “deep learning”, “sensor-based”, “vision-based”, “RFID-based” and, “device-free”. Figure 2 shows the PRISMA diagram showing the criteria for the selection of HAR articles. We identified around 1548 articles in the last 10 years period, which were then short-listed to 175 articles based on three major assessment criteria: AI models used, target application domain, and data acquisition devices which are the three main pillars of the proposed review. In the proposed review we have formed two clusters of attributes based on three major assessment criteria. Cluster 1 includes 7 HAR devices and applications-based attributes, and cluster 2 includes 7 AI attributes. HAR devices and application-based attributes are: data source, #activities, datasets, subjects, scenarios, total #actions and performance evaluation, while the AI attributes includes: #features, feature extraction, ML/DL model, architecture, metrics, validation and hyperparameters/optimizer/loss function.

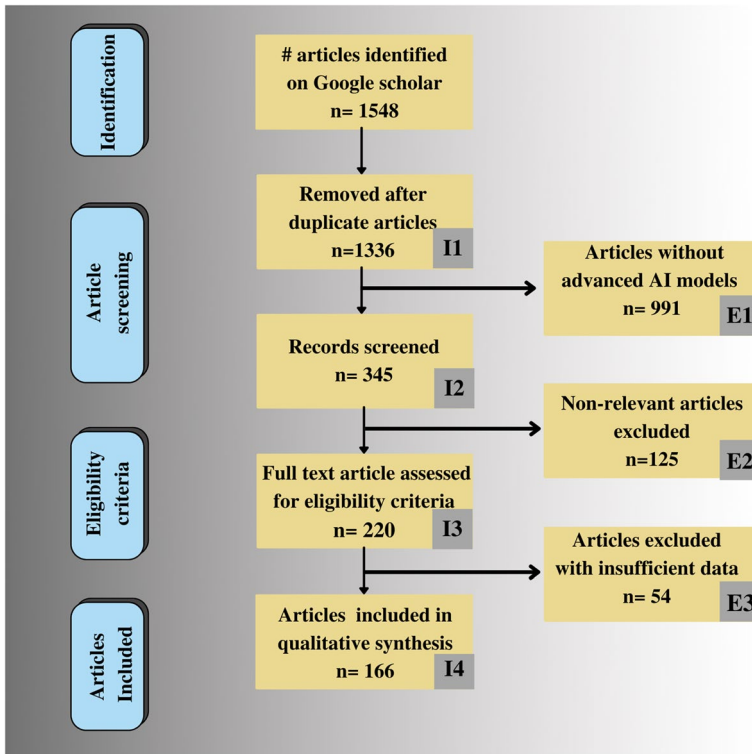


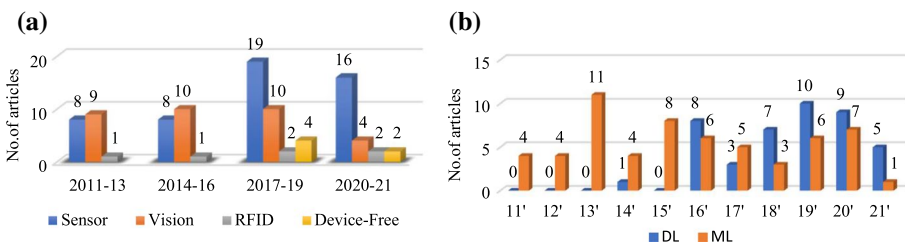
Fig. 2 PRISMA model for the study selection

The description of HAR devices and applications-based attributes is given in Sect. 3.2. Further, the Tables A.1, A.2, A.3 and A.4 of "Appendix 1" illustrate these attributes for various studies considered in the proposed review. The cluster 2's AI attributes are discussed in Sect. 4.2 and Table 3, 4, 5 and 6 illustrate the insight about AI models adapted by researchers in their HAR model. Apart from three major criteria, three exclusion, and four inclusion criteria were also followed in research articles selection. Excluded (1) articles with traditional and older AI techniques, (2) non-relevant articles, and (3) articles with insufficient data. These exclusion criteria consisted of 991, 125, and 54 articles (marked as E1, E2, and E3 in PRISMA flowchart) that lead to the finalization of the 175 articles. Included (1) non-redundant articles, (2) articles with the detailed screening of abstract and conclusion, (3) articles based on eligibility criteria assessment which includes advanced AI techniques, target domain, and device-type, and (4) article's qualitative synthesis including impact factor of journal, and author's contribution in HAR domain; (marked as I1, I2, I3, and I4 in PRISMA flowchart).

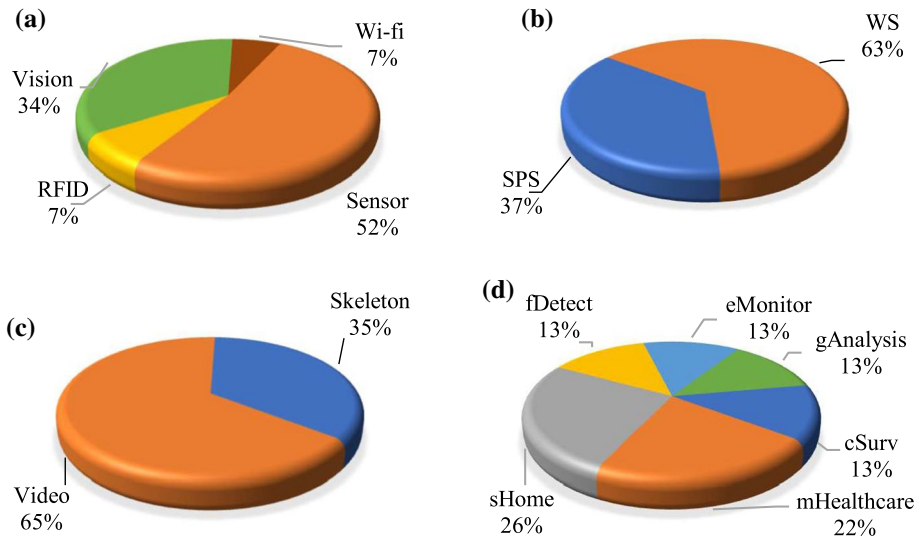
In the proposed review, we performed a rigorous analysis of the HAR framework in terms of AI techniques, device types, and application domain. One of the major observations of the proposed study is the existence of a mutual relationship among HAR device types and AI techniques. First, the analysis on HAR devices is presented in Fig. 3a which is based on the articles considered between the periods of 2011 to 2021. It shows the *changing pattern of HAR devices* over time. Secondly, the growth of ML and DL techniques is presented in Fig. 3b which shows that the HAR is trending towards the use of DL-based techniques. The HAR devices distribution is elaborated more in Fig. 4a, in Fig. 4b we have shown the further categorization of sensor-based HAR into the wearable sensor (WS) and smartphone sensor (SPS). Figure 4c shows the division of vision-based HAR into video and skeleton-based models. Further, Fig. 4d shows the types of HAR application domains.

**Observation 1** In Fig. 3a, according to the device-wise analysis vision-based HAR was popular between the period 2011–2016. But from the year 2017 sensor-based models' growth is more prominent and this is the same time period when DL techniques entered the HAR domain (Fig. 3b). In the period 2017–2021, Wi-Fi devices evolved as one of the data sources for gathering activity.

**Observation 2** Figure 3b shows the year-wise distribution of articles published using ML and DL techniques. The key observation is the transition of AI techniques from ML to DL. From the year 2011–2016, the HAR models with ML framework were popular, while the



**Fig. 3** **a** Changing pattern of HAR devices over time, **b** distribution of machine learning (ML) and deep learning (DL) articles in last decade



**Fig. 4** **a** Types of HAR devices, **b** sensor-based devices, **c** vision-based devices, **d** HAR applications. WS: wearable sensors, SPS: smartphone sensor, sHome: smart home, mHealthcare: health care monitoring, cSurv: crowd surveillance, fDetect: fall detection, eMonitor: exercise monitoring, gAnalysis: gait analysis

HAR models using DL techniques started to evolve from the year 2014. In the last 3 years, this growth has increased significantly. Therefore, after analysing graphs of Fig. 3a, b thoroughly, we can say that the HAR devices are evolving, as the trend is shifting towards the DL framework. This combined analysis verifies our claim of the existence of the mutual relationship between AI and device types.

Devices used in the HAR paradigm are the premier component of HAR by which HAR can be classified. We observed a total of 9 review articles arranged in chronological order (see Table 1). These reviews focused mainly on three sets of devices such as sensor-based (marked in light shade color) (Carvalho and Sofia 2020; Lima et al. 2019; Wang et al. 2016a, 2019b; Lara and Labrador 2013; Hx et al. 2017; Demrozi et al. 2020; Crasto et al. 2019; De-La-Hoz-Franco et al. 2018) or vision-based (marked with dark shade color) (Beddiar et al. 2020; Dhiman Chhavi 2019; Ke et al. 2013; Obaida and Saraei 2017; Popoola and Wang 2012), device-free HAR (Hussain et al. 2020). Table 1 summarizes the nine articles based on the focus area, keywords, number of keywords, research period, and #citations. Note that sensor-based HAR captures activity signals using ambient and embedded sensors, vision-based HAR involves 3-dimensional (3D) activity data gathering using a 3D camera or depth camera. In device-free HAR, activity data is captured using Wi-fi transmitter–receiver units.

**Table 1** Nine review articles published between 2013 and 2020

S. no.	Citation	Year	Focus area	Keywords	#K*	Period	#CIs <sup>§</sup>
1	Wang et al. (2016a)	2020	Device-free HAR	HAR, gesture recognition, motion detection, Device-free, dense sensing, IoT, RFID, human object interaction	7	2011–2017	212
2	Wang et al. (2020)	2020	Sensor-based HAR	Mobile sensing, human-behaviour, human behaviour inference, cloud-edge computing, activity recognition, context awareness	6	2007–2018	136
3	Demrozi et al. (2019b)	2020	Sensor-based HAR	HAR, DL, ML, available datasets, accelerometer, sensors	6	2015–2019	219
4	Beddiar et al. (2020)	2020	Vision-based	HAR, behaviour understanding, Action representation, action detection, computer vision	4	2010–2019	237
5	Dhirman Chhavi (2019)	2019	Vision-based recognition	Two-dimension anomaly detection, three-dimensional anomaly detection, crowd anomaly, skeleton based fall detection, AAL	5	2006–2018	226
6	Lima et al. (2019)	2019	Sensor-based HAR	HAR, smartphones, feature extraction, inertial sensors	4	2006–2017	149
7	Lara and Labrador (2013)	2018	Sensor-based HAR	AAL, HAR, ADL, activity recognition system (ARS), dataset	6	2003–2017	134
8	Hx et al. (2017)	2017	Sensor-based HAR	DL, activity recognition, pervasive computing, pattern recognition	4	2011–2017	80
9	Ke et al. (2013)	2013	Video-based activity recognition	HAR, segmentation, feature representation, health monitoring, security surveillance, human computer interface	5	2003–2012	145

#K keywords, §CIs #citations, AAL ambient assistive living, ADL activity of daily living



### 3 HAR process, HAR devices, and HAR applications in AI framework

The objective of developing HAR models is to provide information about human actions which helps in analyzing the behavior of a person in a real environment. It allows computer-based applications to help users in performing tasks and to improve their lifestyle such as remote care to the elderly living alone, and posture monitoring during exercise. This section presents about HAR framework that includes HAR stages, HAR devices, and target application domains.

#### 3.1 HAR process

*There are four main stages in the HAR process: data acquisition, pre-processing, model training, and performance evaluation (Figure S.1(a) in supporting document).* In *stage 1*, depending on the target application, a HAR device is selected. For example, in surveillance application involving multiple persons, the HAR device for data collection is the camera. Similarly, for applications where a person's daily activity monitoring is involved, the data acquisition source is sensor preferably. One can use a camera also, but it breaches the user's privacy and needs high computational cost. Table 2 illustrates the variation in HAR devices according to the application domains. It elaborates the description of diverse HAR applications in terms of various data sources and AI techniques. Note that sometimes the acquired data suffer from noise or other unwanted signals, and therefore offers challenges in post-processing AI-based systems. Thus, it is very important to have a robust feature extraction system with a robust network for better prediction. In *stage 2*, data cleaning is performed, which involves low-pass or high-pass filters for noise suppression or image enhancement (Suri 2013; Sudeep et al. 2016). This data undergoes regional and boundary segmentation (Multi Modality State-of-the-Art Medical Image Segmentation and 2011; Suri et al. 2002; Suri 2001). Our group has published several dedicated monographs on segmentation paradigms and are available as ready reference (Suri 2004, 2005; El-Baz and Jiang 2016; El-Baz and Suri JS 2019). This segmented data can now be used for model training. *Stage 3* involves the training of HAR model using ML or DL techniques. When using hand-crafted features, one can use ML-based techniques (Maniruzzaman et al. 2017). For automated feature extraction, one can use the DL framework. Apart from automatic feature learning, DL offers knowledge reusability by providing transfer learning models, exploration of huge datasets (Biswas et al. 2018), and hybrid DL models usage which allows spatial as well as temporal features identification and learning. After *stage 3*, the HAR model is ready to be used for an application or prediction. *Stage 4* is the most challenging part since the model is applied to the real data, whose behavior varies depending on physical factors like age, physique, and an approach for performing a task. An HAR model is efficient if its performance is independent of physical factors.

#### 3.2 HAR devices

The HAR device type depends on the target application. *Figure S.1(b) (Supporting document)* presents the different sources for activity data: sensors, video cameras, RFID systems, and Wi-Fi devices.

**Table 2** Multifaceted HAR application with varying HAR devices and AI

App	Application	HAR device	Activity-type	AI model	Architecture	*CIT
eSurv	Crowd surveillance	Subway camera video footage	Group	DL	Laplacian eigenmap feature extraction and k-means clustering-based recognition	Thida et al. (2013)
mHealthcare	Health monitoring	Video data from mobile clips	ADL	DL	Transfer learning-based model using VGG-16 and InceptionV3	Deep and Zheng (2019)
		Smartphone sensor	ADL	DL	CNN (two variants CNN-2 and CNN-7)	Zhu et al. (2019)
		Smartphone sensor	ADL	DL	Hierarchical model-based on DNN	Fazli et al. (2021)
		On-body sensor (Watch and shoe)	ADL	DL	CapSense (a CNN capsule n/w)	Pham et al. (2020)
sHome	Smart home/Smart cities	RFID data collection in an actual trauma room	Single person	DL	CNN model based on data collected from passive RFID tags for trauma resuscitation	Li et al. (2016)
		Smartphone sensor	ADL	DL	CNN (two variants CNN-2 and CNN-7)	Zhu et al. (2019)
		Depth sensor	ADL	DL	3DCNN (Color-skl-MHI and RJI) for elderly care in smart home	Phyo et al. (2019)
		Wireless sensor	AAL	ML	Child activity monitoring based HAR model	Nam and Park (2013)
fDetect	Fall detection	Video	Daily routine	ML	Disabled care HAR model	Jalal et al. (2012)
		Wireless sensor	Forget and Repeat (AAL)	DL	RNN based smart home HAR model for dementia suffering patients	Arifoglu and Bouchachia (2017)
		Smartphone accelerometer	ADL	ML and DL	AdaBoost-HC, AdaBoost-CNN, SVM-CNN	Ferrari et al. (2020)

Table 2 (continued)

App	Application	HAR device	Activity-type	AI model	Architecture	*CIT
eMonitor	Exercise monitoring	Weizmann, KTH with ADL, Ballet: from DVD	ADL and Ballet dance moves	ML	SVM-KNN with PCA	Vishwakarma and Singh (2017)
gAnalysis	Gait analysis	Free weight exercise data recorded with RFID tags SPHERE, DGD: DAI Gait	Free weight exercise Gait pattern	RF ML	FEMO with Doppler shift profile JMh feature and BagofK-eyPoses recognition	Ding et al. (2015) Chaaroui (2015)

\*CIT citation, RF conventional RF profiling, ML machine learning and DL deep learning, ADL activities of daily living, FEMO free weight exercise monitoring, Color-skl-MHI color skeleton motion history images, and RJI relative joint image

**Sensors** The sensors-based approaches can be categorized into wearable sensors and device sensors. In wearable sensor-based approach, a body-worn sensor module is designed which includes inertial sensors, environmental sensors units (Pham et al. 2017, 2020; Hsu et al. 2018; Xia et al. 2020; Murad and Pyun 2017; Saha et al. 2020; Tao et al. 2016a, b; Cook et al. 2013; Zhou et al. 2020; Wang et al. 2016b; Attal et al. 2015; Chen et al. 2021; Fullerton et al. 2017; Khalifa et al. 2018; Tian et al. 2019). Sometimes the wearable sensor devices can be stressful for the user, therefore the solution is the use of smart-device sensors. In device sensor approach data is captured using smartphone inertial sensors (Zhu et al. 2019; Yao et al. 2018; Wang et al. 2016a, 2019b; Zhou et al. 2020; Li et al. 2019; Civitarese et al. 2019; Chen and Shen 2017; Garcia-Gonzalez et al. 2020; Sundaramoorthy and Gudur 2018; Gouineua et al. 2018; Lawal and Bano 2019; Bashar et al. 2020). The most commonly used sensor for HAR is accelerometer and gyroscope. Table A.1 of “Appendix 1”, shows the types of data acquisition devices, activity classes, and scenarios in earlier sensor-based HAR models.

**Video camera** It can be further classified into two types: 3D camera and depth camera. 3D camera-based HAR models uses closed-circuit television (CCTV) cameras in the user’s environment for monitoring the actions performed by the user. Usually, the monitoring task is performed by humans or some innovative recognition model. Numerous HAR models were proposed by researchers, which can process and evaluate the activity video or image data and recognize the performed activities (Wang et al. 2018; Feichtenhofer et al. 2018, 2017, 2016; Diba et al. 2016, 2020; Yan et al. 2018; Chong and Tay 2017). The accuracy of activity recognition of 3D camera data depends on physical factors such as lighting and background color. The solution to this issue can be provided by using a depth camera (like Kinect). The Kinect camera consists of different data streams such as depth, RGB, and audio. Depth stream captures body joint coordinates, and based on joint coordinates, a skeleton-based HAR model can be developed. The skeleton-based HAR models have applications in domains that involve posture recognition (Liu et al. 2020; Abobakr et al. 2018; Akagündüz et al. 2016). Table A.2 of “Appendix 1” provides an overview of earlier vision-based HAR models. Apart from 3D and depth cameras, one can use thermal cameras but it can be expensive.

**RFID tags and readers** By installing RFID passive tags in close proximity of the user, the activity data can be collected using RFID readers. As compared to active RFID tags, passive tags have more operational life as they do not need a separate battery. Rather it uses the reader’s energy and converts it into an electrical signal for operating its circuitry. But the range of active tags is more than passive tags. They both can be used for HAR models (Du et al. 2019; Ding et al. 2015; Li et al. 2016; Yao et al. 2018; Zhang et al. 2011; Xia et al. 2012; Fan et al. 2019). The further description of existing RFID-based HAR models is provided in Table A.3 of “Appendix 1”.

**Wi-Fi device:** In the last 5 years, the device-free HAR has gained popularity. Researchers have explored the possibility of capturing activity signals using Wi-Fi devices. Channel state information (CSI) from the wireless signal is used to acquire activity data. Many models were developed for *fall detection* and *gait recognition* using CSI (Yao et al. 2018; Wang et al. 2019c, d, 2020b; Zou et al. 2019; Yan et al. 2020; Fei et al. 2020). The description of some popular existing Wi-Fi device-based HAR is provided in Table A.4 of “Appendix 1”.

**Summary of challenges in HAR devices** There are almost four types of HAR devices, and researchers have proposed various HAR models with advanced AI techniques. Gradually, the usage of electronic devices for gathering activity data in HAR domain is increasing, but with this growth, the challenges are also evolving: (1) Video camera-based application involves data gathering using a video camera, which results in the invasion of user's privacy. It also requires high power systems to process large data produced by video cameras, (2) In sensors-based HAR models, the use of wearable devices is stressful and inconvenient for the user, therefore smartphone sensors are more preferable. But the use of smartphone and smartwatch is limited to simple activities recognition such as walking, sitting, and going upstairs, (3) In RFID tags and reader-based HAR models, the usage of RFID in activity capturing is limited to indoor only. (4) Wi-Fi-based HAR models are new in the HAR industry, but there are few issues with it. Moreover, it can capture activities performed within the Wi-Fi range but cannot identify the movement in blind spot areas.

### 3.3 HAR applications using AI

In the last decade, researchers have developed various HAR models for different domains. "What type of HAR device is suitable for which application domain and what is the suitable AI methodology" is the biggest question that pops into the mind, once developing the HAR framework. The description of diverse HAR applications with data sources and AI techniques is illustrated in Table 2. It shows the variation in HAR devices and AI techniques depending on the application domain. The pie chart in Fig. 4d shows the distribution of applications based on existing articles. HAR is used in fields like:

- Crowd surveillance (cSurv): Crowd pattern monitoring and detecting panic situations in the crowd.
- Health care monitoring (mHealthcare): Assistive care to ICU patients, Trauma resuscitation.
- Smart home (sHome): Care to elderly or dementia patients and child activity monitoring.
- Fall detection (fDetect): Detection of abnormality in action which results in a person's fall.
- Exercise monitoring (eMonitor): Pose estimation while doing exercise.
- Gait analysis (gAnalysis): Analyze *gait* patterns to monitor health problems.

### 3.4 HAR applications with different activity-types

There is no predefined set of activities, rather the human activity type varies according to the application domain. Figure S.2 (Supporting document) shows the activity type involved in human activity recognition.

**Single person activity** Here the action is performed by a person. Figure S.3 (Supporting document) shows examples of single-person activities (jumping jack, baby crawling, punching the boxing bag, and handstand walking). Single person action can be divided into the following categories:

- *Behavior*: The goal of behavior recognition is to recognize a person's behavior from activity data, and it is useful in monitoring applications: dementia patient & children behavior (Han et al. 2014; Nam and Park 2013; Arifoglu and Bouchachia 2017).
- *Gestures*: It has application in sign language recognition for differently-abled persons. Wearable sensor-based HAR models are more suitable (Sreekanth and Narayanan 2017; Ohn-Bar and Trivedi 2014; Xie et al. 2018; Kasnesis et al. 2017; Zhu and Sheng 2012).
- *Activity of daily living (ADL) and Ambient assistive living (AAL)*: ADL activities are performed in an indoor environment cooking, sleeping, and sitting. In smart home, ADL monitoring for dementia patients can be performed using wireless sensor-based HAR models (Nguyen et al. 2017; Sung et al. 2012) or RFID tags based HAR models (Ke et al. 2013; Oguntala et al. 2019; Raad et al. 2018; Ronao and Cho 2016). AAL-based models help elderly and disabled people by providing remote care, medication reminder, and management (Rashidi and Mihailidis 2013). CCTV cameras are an ideal choice but they have privacy issues (Shivendra shivani and Agarwal 2018). Therefore, sensor or RFID-based HAR models (Parada et al. 2016; Adame et al. 2018) or wearable sensor-based models are more suitable (Azkune and Almeida 2018; Ehatisham-UI-Haq et al. 2020; Magherini et al. 2013).

**Multiple person activity** The action is performed by a group of persons. Multiple person movement is illustrated in Figure S.4 (Supporting document), depicts the normal human movement on a pedestrian pathway and anomalous activity of cyclist and truck in a pedestrian pathway. It can belong to the following categories.

- *Interaction*: There are human–object (cooking, reading a book) (Kim et al. 2019; Koppula et al. 2013; Ni et al. 2013; Xu et al. 2017) and human–human (handshake) activities (Weng et al. 2021). A human–object interaction-based free weight exercise monitoring (FEMO) model using RFID devices that monitors exercise by installing a tag on dumbbells (Ding et al. 2015).
- *Group*: It involves monitoring people's count in an indoor environment like a museum or crowd pattern monitoring (Chong and Tay 2017; Xu et al. 2013). To check the number of people in an area, we can use Wi-Fi units. Received signal strength can be used for counting people as it is user-sensitive.

Observation 3: Vision-based HAR has broad application domains, but they have limitations like privacy and the need for more resources (such as GPUs). These issues can be overcome with sensor-based HAR but their applications domain is currently limited to single-person activity monitoring.

## 4 Core of the HAR system design: emerging AI

The foremost goal of HAR is to predict the movement or action of a person based on the action data collected from a data acquisition device. These movements include activities like walking, exercising, and cooking. It is challenging to predict movements, as it involves huge amounts of unlabelled sensor data, and video data which suffer from conditions like

lights, background noise, and scale variation. To overcome these challenges AI framework offers numerous ML, and DL techniques.

#### 4.1 Artificial intelligence models in HAR

**ML architectures:** ML is a subset of AI, which aims at developing an intelligent model which involves the extraction of unique features, that helps in recognizing patterns in the input data (Maniruzzaman et al. 2018). There are two types of ML approaches: supervised and unsupervised. In supervised approach, a mathematical model is created based on the relationship between raw input data and output data. The idea behind the unsupervised approach is to detect patterns in raw input data without prior knowledge of output. Figure S.5 (Supporting document) illustrates the popular ML techniques used in recognizing human actions (Qi et al. 2018; Yao et al. 2019; Multi Modality State-of-the-Art Medical Image Segmentation and 2011). Several applications of ML models in handling different diseases have been developed by our group such as diabetes man(Maniruzzaman et al. 2018) liver cancer (Biswas et al. 2018), thyroid cancer (Rajendra Acharya et al. 2014), ovarian cancer (Acharya et al. 2013a, 2015), prostate (Pareek et al. 2013) breast (Huang et al. 2008), skin (Shrivastava et al. 2016), arrhythmia classification (Martis et al. 2013), and recently in cardiovascular (Acharya et al. 2012; Acharya et al. 2013b). In the last 5 years, the researchers' focus has been shifted to semi-supervised learning where the HAR model is trained on labelled as well as unlabelled data. The semi-supervised approach aims to label unlabelled data using the knowledge gained from the set of labelled data. In a semi-supervised approach, the HAR model is trained on popular labelled datasets and the new users' unlabelled test data and classified into activity classes according to the knowledge gained from training data (Mabrouk et al. 2015; Cardoso and Mendes Moreira 2016).

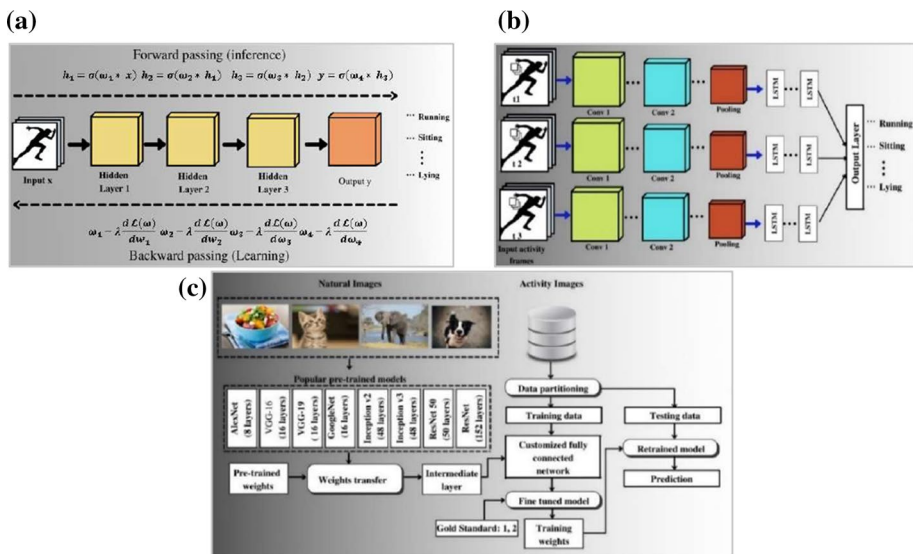
**DL/TL Architectures:** In recent years, DL has become quite popular due to its capability of learning high-level features and its superior performance (Saba et al. 2019; Biswas et al. 2019). The basic idea behind DL is data representation, which enables it to produce optimal features. It learns unknown patterns from raw data without human intervention. The DL techniques used in HAR can be divided into three parts such as deep neural networks (DNN), hybrid deep learning (HDL) models, and transfer learning (TL) based models (Agarwal et al. 2021). (Shown in Figure S.5 of Supporting document) The DNN includes the models like convolutional neural networks (CNN) (Deep and Zheng 2019; Liu et al. 2020; Zeng et al. 2014), recurrent neural networks (RNN) (Murad and Pyun 2017) and RNN variants which include long short-term memory (LSTM) and gated recurrent unit (GRU) (Zhu et al. 2019; Du et al. 2019; Fazli et al. 2021). In hybrid HAR models, the combination of CNN and RNN models is trained on spatio-temporal data. Researchers have proposed various hybrid models in the last 5 years, such as DeepSense (Yao et al. 2017) and DeepConvLSTM (Wang et al. 2019a). Apart from hybrid AI models, there are various transfer learning-based HAR models which involves pre-trained DL architectures like ResNet-50, Inceptionv3, VGG-16 (Feichtenhofer et al. 2018; Newell Alejandro 2016; Crasto et al. 2019; Tran et al. 2019; Feichtenhofer and Ai 2019). However, the role of TL in sensor-based HAR is still evolving (Deep and Zheng 2019).

Figure 5a depicts a representative CNN architecture for HAR, which shows the two convolution layers followed by a pooling layer for feature extraction for the activity image, leading to dimensionality reduction. This is then followed by a fully connected (FC) layer for iterative weight computations and a softmax layer for binary or granular decision making. After that, the input image is classified into an activity class. Figure 5b presents the

representative TL-based HAR model, which includes pretrained models such as VGG-16, inception V3, and ResNet. The pre-trained model is trained on a large dataset of natural images such as man, cat, dog, and food. These pre-trained weights are applied to the training data of the sequence of images using an intermediate layer. It forms the customized fully connected layer. Further, the training weights are fine-tuned using the optimizer function. Next the retrained model is applied to testing data for the classification of the activity into an activity class.

Miniaturized mobile devices are handy to use and offer a set of physiological sensors that can be used for capturing activity signals. But the problem is the complex structure and strong inner correlation in captured data. The deep learning models which are the combination of both CNN and RNN offer benefits to explore this complex data and identify detailed features for activity recognition. One such model offered by Ordonez et al. was DeepConvLSTM (Ordóñez and Roggen 2016), where CNN works as feature extractor and represent the sensor input data as feature maps, and LSTM layer explores the temporal dynamics of feature maps. Yao et al. have proposed similar model named as DeepSense in which two convolution layers (individual and merge conv layers) and stacked GRU layers were used as main building blocks (Yao et al. 2017). Figure 5c shows the representative hybrid HAR model with CNN-LSTM frameworks.

**Loss function** DL model learns by means of loss function. It evaluates how well an algorithm models the applied data. If it deviates largely from actual output, the value of the loss function will be very large. The loss function with the help of optimization function learns gradually to reduce the prediction error. Mostly used loss functions in HAR models are mean squared loss and cross-entropy (Janocha and Czarnecki 2016; Wang et al. 2020a).



**Fig. 5 a** CNN model for HAR (where  $\omega_{(*)}$ : weights of hidden layers,  $\sigma(*)$ : activation function,  $\lambda$ : learning rate,  $*$ : convolutional operation,  $\mathcal{L}(\omega)$  is the loss function), **b** TL-based model for HAR, and **c** hybrid HAR model (CNN-LSTM)



- *Mean absolute error* ( $\delta$ ): is calculated as the average sum of absolute differences between predicted ( $y_i$ ) and actual ( $\hat{y}_i$ ) output.  $N$  is the number of training samples

$$\delta = \frac{\sum_{i=0}^N |y_i - \hat{y}_i|}{N}$$

- *Mean squared error* ( $\epsilon$ ): is calculated as the average of the squared difference between the predicted ( $\hat{y}_i$ ) and actual output ( $y_i$ ).  $N$  is the number of training samples

$$\epsilon = \frac{\sum_{i=0}^N (y_i - \hat{y}_i)^2}{N}$$

- *Cross-entropy loss* ( $\eta$ ): evaluates the performance of a model whose output probability ranges between 0 and 1. The loss increases if predicted probability ( $y_i$ ) diverges from actual output ( $\hat{y}_i$ ).
- *Binary cross-entropy loss*: predict the probability between two activity classes.

$$\eta = -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

- *Multiclass cross-entropy loss*: Multi-class CEL is the generalization of binary CEL where each class is assigned a unique integer value range between 0 to  $n - 1$  ( $n$  is a number of classes).

$$\rho = - \sum_{n=0}^N y_{i,n} \log(\hat{y}_{i,n})$$

- *Kullback Lieblar-divergence* (KL-divergence): is a measure of how a probability distribution diverges from another distribution. For the probability distribution of  $P(x)$  and  $Q(x)$ , KL-divergence is defined as the logarithmic difference between  $P(x)$  and  $Q(x)$  with respect to  $P(x)$ .

$$\text{KL}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

### Hyper-parameters and optimization

- *Drop-out rate*: regularization technique where few neurons are dropped to avoid over-fitting.
- *Learning rate*: it defines how fast parameters are updated in a network.
- *Momentum*: it helps in the next step direction based on knowledge gained in previous steps.
- *Number of hidden layers*: number of hidden layers between input and output layers.

**Optimization** It is a method used for changing the parameters of neural networks. DL provides a wide range of optimizers: gradient descent (GD), stochastic gradient descent (SGD), RMSprop, and Adam optimizers. GD is a first-order optimization that relies on the first-order derivative of the loss function. SGD is the variant of GD, which involves frequent variation in a model's parameter. It computes the loss for each training sample and alters the model's parameters. Further, the RMSprop optimizer lies in the domain of adaptive learning. RMSprop deals with the vanishing/exploding gradient issue by using

a moving average of squared gradients to normalize the gradient. The most powerful optimizer is the Adam optimizer which has the strength of momentum of GD to hold the gained knowledge of updates, adaptive learning of RMSprop optimizer, offers two new hyper-parameters beta and beta 2 (Soydaner 2020; Sun et al. (2020).

**Validation** The most common validation strategies are K-fold cross validation and leave one subject out (LOSO). In k-fold, the k-onefold is used for training and the remaining is used for validation. A similar pattern is followed in k-fold variants such as twofold, three-fold, and tenfold cross-validation. In LOSO, out of whole dataset, the data of one subject is kept for validation and the rest is used for training.

## 4.2 AI models adapting HAR devices

There are various HAR devices for capturing human activity signals. The goal of HAR devices is to capture activity signals with minimal distortion. For providing deeper insight into the existing HAR models, we have identified seven AI attributes and used tabular representation for better understanding. It consists of attributes such as #features, feature extraction, AI model architecture, metrics, validation, hyper-parameters/optimizer/loss function. For in-depth description of recent HAR models between 2019 and 2021, we have made four tables for each HAR device: Table 3 (sensor), Table 4 (vision), Table 5 (RFID), and Table 6 (device-free).

In Table 3, we have provided insight into AI techniques adopted in sensor-based HAR models in the last two years. Apart from recent sensor-based HAR models, knowledge about previous sensor-based HAR models published between 2011–2018 is provided in Table S.1 (Supporting document) (Zhu et al. 2019; Ding et al. 2019; Yao et al. 2017, 2019; Hsu et al. 2018; Murad and Pyun 2017; Sundaramoorthy and Gudur 2018; Lawal and Bano 2019). The sensor-based HAR is more dominated by DL techniques especially CNN or the CNN's combination with RNN or its variants. In sensor-based HAR the most used hyper-parameters are learning rate, batch size, #layers, and drop out. Adam optimizer, cross-entropy loss, and k-fold validation are dominant in sensor-based HAR. For example, Table 3's (R2, C4) presents the 3D CNN-based HAR model which includes 3 convolutional layers of size (32, 64,128) followed by a pooling layer, then an FC layer of size (128) and a softmax layer. Entry (R2, C6) illustrate the validation strategy (10% data was used for validation) and entry (R2, C7) illustrates the hyperparameters (i.e., LR=0.001, batch size=50) and selected optimiser (Adam) for performance fine-tuning. Table 4 illustrates the AI framework in vision-based HAR models published in recent 2 years. Further, description of earlier vision-based HAR models published between 2011–2018 are provided in Table S.2 (Supporting document) (Qi et al. 2018; Wang et al. 2018; Thida et al. 2013; Feichtenhofer et al. 2018, 2017; Simonyan and Zisserman 2014; Newell Alejandro 2016; Diba et al. 2016; Xia et al. 2012; Vishwakarma and Singh 2017; Chaaraoui 2015). Initial vision-based HAR models were dominated by ML algorithms such as support vector machine (SVM), k-means clustering with principal component analysis (PCA)-based feature extraction. In the last few years, researchers have shifted to DL paradigm and the most dominant DL techniques such as multi-dimensional CNN, LSTM, and a combination of both. In video camera-based HAR models, the incoming data is video stream which needs more resources and processing time. This issue gives rise to the usage of TL in vision-based HAR approaches. The hyper-parameters used in vision-based HAR are drop-out rate, learning rate, weight decay, and batch normalization. The mean square

**Table 3** Sensor-based HAR models

C1	C2	C3	C4	C5	C6	C7	C8
# Features	Feature extraction	ML or DL model	Architecture	Metrics	Validation	Hyper-parameters/optimizer/loss function	CIT*
R1	6/Time domain	SVM	SVM classifier for different kernels (polynomial, Radial basis function and linear)	F1-score, accuracy	tenfold	C, $\gamma$ & degree in grid search	Garcia-Gonzalez et al. (2020)
R2	Spatial features	CNN	C (32) – C (64) – C (128) – P – C (128) – P – C (128) – P – FC (128) – SM	Accuracy	10% data for validation	LR: 0.001, BS: 50/Adam	Wang et al. (2019a)
R3	Frequency domain	CNN	3C with MP and dropout, 2 FC with dropout and SM	F1-score, Precision, recall	CV	LR: 0.01, DO/Adam	Lawal and Bano (2020)
R4	Time domain	CNN-RNN with attention mechanism	TRASEND: C1-merge layer-tem-poral information extractor using a 8-headed self-attention mechanism RNN, o/p layer	F1-score	Leave one user out and CV	LR: {0.001, 0.0001, 0.00001}/Adam/Cross Entropy	Buffelli and Vandin (2020)

Table 3 (continued)

C1	C2	C3	C4	C5	C6	C7	C8
# Features	Feature extraction	ML or DL model	Architecture	Metrics	Validation	Hyper-parameters/ optimizer/loss function	CIT*
R5	Spatial, Temporal	Automatic	LSTM-CNN	F1-score, accuracy	-	LR: 0.001/Adam/ Cross entropy	Xia et al. (2020)
R6	18/Time & Frequency domain	Hand-crafted	AdaBoost, AdaBoost-CNN, CNN-SVM	Accuracy	Sub-out validation	Experiment with and without personalization similarity	Ferrari et al. (2020)
R7	225 sensory features	Automatic	DNN	Accuracy, F1-score, Specificity, Sensitivity	5% training data is used	No. of layers, no. of nodes per layer, appropriate regularization function	Fazli et al. (2021)
R8	Time domain	Automatic	CNN-CapsNet architecture	Precision, recall	tenfold CV	mini batches:64,LR: 0.01, DO/SGD	Pham et al. (2020)

CV cross validation, *LOSO* leave one subject out, *C* convolution, *P* pooling, *AP* average pooling, *MP* max pooling, *FC* fully connected, *SM* softmax, *BN* batch normalization layer, *LR* learning rate, *DO* dropout, *BS* batch size, *SGD* stochastic gradient descent, *concat* concatenation, *Spec.* specificity, *Sens* sensitivity, *TL* transfer learning, *CIT* citations

**Table 4** Vision-based HAR models

C1	C2	C3	C4	C5	C6	C7	C8
# Features	Feature extraction	ML/DL model	Architecture	Metrics	Validation	Hyper-parameters/optimizer/loss function	CIT*
R1	T. domain Hand-crafted	3DCNN on Color-skl-MHI and RJI	I/p layer with skeletal joints, Color-skl-MHI followed by 3D-DCNN, RJI followed by 3D-DCNN, decision fusion, o/p	Accuracy	Cross validation	DO ratios for the three hidden layers (0.1%, 0.2%,0.3%/SGD	Phyo et al. (2019)
R2	Spatio temporal Automatic	VGG-16, VGG-19, inception v3	224×224 is image is input and features from fc1 layer are extracted which gives 4096-dimensional vector for per image	Accuracy, precision, recall, F1-score	10% data is used for validation	All 3 CNNs trained on imageNet then trained on Weizmann reusing same weights	Deep and Zheng (2019)
R3	Spatio temporal Automatic	ResNet-50	C1, MP, C2-C5, AP, FC (2048), SM	Accuracy	Evaluates UCF-101 and HMDB-51	BS- 128, DO, LR: $10^{-4}$ and $10^{-5}$ /SGD	Feichtenhofer et al. (2017)
R4	Spatio temporal Automatic	ResNet-50	Raw clip i/p C1- P-C2- C3- C4- C5- GAP- FC- No. of classes. Pre-train on Kinetics-400, Kinetics-600 and kinetics-700	mAPS, GFLOPS	Evaluate model performance on AVA dataset	LR, WD: $10^{-7}$ , Batch normalization/SGD	Feichtenhofer and Ai (2019)

Table 4 (continued)

C1	C2	C3	C4	C5	C6	C7	C8	
# Features	Feature extraction	ML/DL model	Architecture	Metrics	Validation	Hyper-parameters/ optimizer/loss function	CIT*	
R5	Spatio temporal	Automatic	MERS model with ResNeXt-101	MERS: Train using flow, freeze weights, train with RGB using MSE loss. MARS: Train using privileged flow n/w, freeze weights, use RGB frames during test phase	top-1 mean accuracy	Kinetics 40: 20 k, MiniKinetics: 5 k	WD = 0.0005, LR = 0.1, momentum = 0.9 and LR = 0.1 for 64f-clips/SGD/Cross entropy	Crasto et al. (2019)
R6	Spatio temporal	Automatic	HATnet based on ResNet-50 and STCnet	2D ConvNets: to extract spatial structure, 3DCConv: to deal with interaction in frames. Both 2D and 3D use ResNet-50	Top-1 mAPS	Kinetics 400 and 600	Fine tune on UCF-101 & HMDB-51/Cross entropy	Diba et al. (2020)

Table 4 (continued)

C1	C2	C3	C4	C5	C6	C7	C8
# Features	Feature extraction	ML/DL model	Architecture	Metrics	Validation	Hyper-parameters/ optimizer/loss function	CIT*
R7	Spatio-temporal	Automatic	Video frames i/p, C1, C2x, C3x, C4x, C5x, FC, o/p. Replace all residual block with STM block (1×1 2D conv, followed by CMM and CSTM blocks, then 1×1 2D Conv)	top-1, top-5 accu- racy	Kinetics 400: 19,095	LR=0.01, LR=0.001 for 25 epochs, momentum=0.9, WD=2.5 e <sup>-4</sup> / SGD	Jiang et al. (2019)

*T* time, *F* frequency, *CV* cross validation, *LOSO* leave one subject out, *C* convolution, *P* pooling, *AP* average pooling, *MP* max pooling, *FC* fully connected, *SM* softmax, *BN* batch normalization layer, *LR* learning rate, *DO* dropout, *BS* batch size, *SGD* stochastic gradient descent, *mAP<sub>s</sub>* mean average precision, *GFLOP* giga floating point operations per second, *Spec* specificity, *Sens* sensitivity, *AUC* area under curve, *EEER* equal error rate, *TL* transfer Learning

\*CIT citations

Table 5 RFD-based HAR models

C1	C2	C3	C4	C5	C6	C7	C8
# features/ F, Fisher	Feature extraction	ML/DL model	Architecture	Metrics	Validation	Hyper-parameters/ Optimizer Loss Function	CIT*
R1	84 features for F-statistics, Relief- F, Fisher	Hand-crafted Canonical Cor- relation Analysis (CCA)	Divide RSSI stream into segments, CCA for extracting features by com- puting canonical correlation for each feature pair, activity specific dictionary is formed. Sparse coding and dictio- nary is updated sequentially using K-SVD	F1-score	One sub out validation strategy	-	Yao et al. (2018)
R2	Frequency	Automatic LSTM	I/p layer, two hidden layers and o/p layer	Precision, accuracy	-	Time-step, neuron in hidden layers/ Adam/Cross entropy	Du et al. (2019)
R3	Frequency	Hand-crafted Multi variate Gauss- ian Approach	ADL activity data gathering, score each activity with gaussian pdf. Human Activity Recognition Based on Maximum likelihood Estima- tion, Activity classification	Accuracy, precision, recall, F1-score, root mean square error (RMSE)	-	-	Oguntala et al. (2019)

\*CIT citations, DTW dynamic time wrapping, DO dropout, LR learning rate, SVM support vector machine, LSTM long short-term memory



**Table 6** Device-free HAR models

C1	C2	C3	C4	C5	C6	C7	C8
# Features	Featureextraction	ML/DL model	Architecture	Metrics	Validation	Hyper-parameters/opti- mizer loss function	CIT*
R1	Frequency	Hand-crafted	ELM	Accuracy	Tested the performance on 1100 samples	No. of hidden layer neurons (after 400 becomes stable), different users, impact of total no. of samples	Yan et al. (2020)
R2	Frequency CSI	Hand-crafted	DTW: by comparing similarity b/w waveforms	Accuracy	Recognition of gaits using MARS	Impact of nearby people, test for system delay	Fei et al. (2020)
R3	CSI (time & amplitude)	Automatic	LSTM	FPR, precision, recall, F1-score		DO, Layer size, recognition method	Wang et al. (2019d)

\*CIT citations, DTW dynamic time wrapping, FPR false positive rate, DO dropout, ELM extreme learning machine, CSI channel state information

loss and cross-entropy loss are the most used loss functions, while RMSProp and SGD are the most dominant optimizers in vision-based HAR. For example, Table 4's (R1, C3) illustrates the description of 3DCNN based HAR model which includes input layer with skeletal joints information split into coloured skeleton motion history images (Color-skl-MHI), and relative joint images (RJI) followed by 3DCNN, then a fusion layer to combine the o/p of both 3DCNN layers and last is the output layer. Table 5 shows the recognition models using RFID devices published in the last 2 years, while details of the earlier RFID-based HAR models are provided in Table S.3 (Supporting document) (Ding et al. 2015; Li et al. 2016; Fan et al. 2017). RFID-based HAR is mostly dominated by ML algorithms like SVM, sparse coding, and dictionary learning. Very few researchers have used DL techniques. Some RFID-based HAR models used traditional approach in which received signal strength indicator (RSSI) is used for data gathering and recognition task is performed by calculating the similarity in dynamic time warping (DTW). Table 6 provides the overview of device-free HAR models where Wi-Fi devices are used for collecting activity data. The recognition approach is similar to RFID-based HAR. Further, ML approaches are more dominant than DL.

**Impact of DL on miniaturized mobile and wireless sensing HAR devices** A visible growth of DL in vision-based HAR devices is observed in terms of existing HAR models mentioned in Table 4 where most of the recent work is done using advanced DL techniques like TL using VGG-16, VGG-19, and ResNet-50. Apart from these TL-based models, there are hybrid models using autoencoders as shown in row R8 of Table 4 which includes CNN, LSTM, and autoencoder-based HAR model for extracting deep features from enormous volumed video datasets. But the impact of advanced DL techniques in sensors-based HAR and device-free HAR is not very powerful. Due to the compact size and versatility of miniaturized and wireless sensing devices, they are progressing to become the next revolution in the HAR framework, and the key to their progress is the emerging DL framework. The data gathered from these devices is unlabelled, complex, and has strong inter-correlation. DL offers (1) advanced algorithms like TL, and unsupervised learning techniques such as generative adversarial networks (GAN) and variational autoencoders (VAE), (2) fast optimization techniques such as SGD, Adam, and (3) dedicated DL libraries like TensorFlow, (Py) Torch, and Theano to handle complex data.

Observation 4: DL techniques are still in an evolving stage. Minimal work has been done using TL in sensor-based HAR models. Most of the approaches are discriminative where supervised learning is used for training HAR models. Generative models like VAE and GAN have evolved in the computer vision domain but they are still new in the HAR domain.

## 5 Performance evaluation in HAR and integration of AI in HAR devices

### 5.1 Performance evaluation

Researchers have adopted different metrics for evaluating the performance of HAR models, and the most popular evaluation metric is accuracy. The most used metrics in sensor-based HAR include accuracy, sensitivity, specificity, and F1-score. The evaluation metrics used in existing vision-based HAR models were accuracy i.e., top-1, top-5, and mean average

precision (mAPS). Metrics used in RFID-based HAR include accuracy, F1-score, recall, and precision. The metrics used in Device-free HAR include F1-score, precision, recall, and accuracy. “Appendix 2” shows the mathematical representations of the performance evaluation metrics used in the HAR framework.

## 5.2 Integration of AI in HAR devices

In the last few years, a significant growth can be seen in the usage of DL in the HAR framework, but there are challenges associated with DL models such as (1) *Overfitting/Underfitting*: When the amount of activity data is limited, the HAR model learns too well during training that it learns the irregularities and random noise as part of data. As a result, it negatively impacts the model’s generalization ability. Underfitting is another negative condition where the HAR model neither models the new data nor generalizes to new unseen data. Both *overfitting* and *underfitting* result in lower performance. By selecting the appropriate optimizer, we can overcome the overfitting condition by tuning the right hyperparameters or by increasing the size of training data, or using k-fold cross validation. The challenge is to select the correct range of hyperparameters that can work well during training and testing protocols and works well when the HAR model is used in real-life applications. (2) *Hardware integration in HAR devices*: In the last 10 years various HAR models with high performance came into the picture, but the question is “how well they can be used in real-environment without integrating specialized hardware like graphics processing units (GPUs), and extra memory”. Therefore, the objective for designing a HAR model is to design a robust and lightweight model which can run in real-environment without the need for specialized hardware. For applications with huge data such as videos, we need GPUs for training the model. Python offers libraries (such as Keras, TensorFlow) for implementing AI framework on a general-purpose CPU processor. For working on GPUs, one needs to explore special libraries for implementing AI models. Sometimes, it may result in specialized hardware integration need in the target application which makes it expensive. Processing power and costs are interrelated i.e., one needs to pay more for extra power.

## 6 Critical discussion

In the proposed review, we made four observations based on the tri-stool of HAR which includes HAR devices, AI techniques, and applications. Based on these observations and challenges highlighted in Sects. 3 and 4, we have made three claims and four recommendations.

### 6.1 Claims based on in-depth analysis of HAR devices, AI, and HAR applications

(i) *Mutual relationship among HAR devices and AI framework*: Our first claim is based on the observation 1 and 2 where we illustrate that the advancement in AI directly affects the growth of HAR devices. In Sect. 2, Fig. 3a presents the growth of HAR devices in the last 10 years. Further, Fig. 3b illustrates the advancement in AI, which shows how researchers

have shifted to the DL paradigm from ML in the last 5 years. Therefore, from observations 1 and 2, we can rationalize that the advancement in AI is resulting in the growth of HAR devices. Most of the earlier HAR models were dependent on cameras or customized wearable sensors data but in the last 5 years more devices like embedded sensors, Wi-Fi devices came into the picture as prominent HAR sources.

(ii) *Growth in HAR devices increases the scope of HAR in various application domains:* Claim 2 is based on observation 3, where we have shown that for the best results how a target application is depending on a HAR device. For applications like crowd monitoring, if we use sensor devices for gathering the activity data it will not be able to give prominent results because sensors are best for single person applications. Similarly, if we use a camera in a smart home environment, it will not be a good choice because cameras invade user's privacy and require a high computational cost.

Therefore, we can conclude that multi-person applications like surveillance video cameras are proven best. However, for single-person monitoring applications smart device sensors are more suitable.

(iii) *HAR devices, AI, and target application domains are three pillars in HAR framework:* From all four observations and claims (1 and 2), we have proved that HAR devices, AI, and application domains are three pillars in the success of a HAR model.

## 6.2 Benchmarking: comparison between different HAR reviews

The objective of the proposed review is to provide a complete and comprehensive review of HAR based on the three pillars i.e., device-type, AI techniques, and application domains. Table 7 provides the benchmarking of the proposed review with existing studies.

## 6.3 A short note on HAR datasets

The narrative review surely needs a special note on types of HAR datasets. (1) *Sensor-based:* Researchers have proposed many popular sensor-based datasets. In Table A.5 ("Appendix 1"), the description of sensor-datasets is illustrated with attributes such as data source, #factors, sensor location, and activity type. It includes wearable sensor-based datasets (Alsheikh et al. 2016; Asteriadis and Daras 2017; Zhang et al. 2012; Chavarriga et al. 2013; Munoz-Organero 2019; Roggen et al. 2010; Qin et al. 2019), as well as smart-device sensor-based datasets (Ravi et al. 2016; Cui and Xu 2013; Weiss et al. 2019; Miu et al. 2015; Reiss and Stricker 2012a, b; Lv et al. 2020; Gani et al. 2019; Stisen et al. 2015; Röcker et al. 2017; Micucci et al. 2017). Apart from datasets mentioned in Table A.5, there are few more datasets worth mentioning such as Kasteren dataset (Kasteren et al. 2011; Chen et al. 2017), which is also very popular. (2) *Vision-based HAR:* Devices for collecting 3D data are CCTV cameras (Koppula and Saxena 2016; Devanne et al. 2015; Zhang and Parker 2016; Li et al. 2010; Duan et al. 2020; Kalfaoglu et al. 2020; Gorelick et al. 2007; Mahadevan et al. 2010), depth cameras (Cippitelli et al. 2016; Gaglio et al. 2015; Neili Boualia and Essoukri Ben Amara 2021; Ding et al. 2016; Cornell Activity Datasets: CAD-60 & CAD-120 2021), and videos from public domains like YouTube and Hollywood movie scenes (Gu et al. 2018; Soomro et al. 2012; Kuehne et al. 2011; Sigurdsson et al. 2016; Kay et al. 2017; Carreira et al. 2018; Goyal et al. 2017). The reason behind using public domain videos is that they have no privacy issue, unlike with cameras. Table A.6 ("Appendix 1") illustrates the description of *vision-based* datasets which includes

data source, #factors, sensor location, and activity type. Apart from datasets mentioned in Table A.6, there are few more publicly available datasets such as MCDS (Magnetic wall chess board video) datasets (Tanberk et al. 2020), NTU-RGBD datasets (Yan et al. 2018; Liu et al. 2016), VIRAT 1.0 (3 hour person vehicle interaction), and VIRAT 2.0 (8 hour surveillance scene of school parking) (Wang and Ji 2014). (3) *RFID-based*: RFID-based HAR is mostly used for smart home applications, where actions performed by the user are monitored by RFID tags. To the best of our knowledge, there is hardly a public dataset available for *RFID-based* HAR. Researchers have developed their own datasets for their respective applications. One such dataset was developed by Ding et al. (Ding et al. 2015) in 2015 which includes data of 10 exercises performed by 15 volunteers for 2 weeks with a total duration of 1543 min. Similarly, Li et al. developed the dataset for trauma resuscitation including the 10 activities and 5 resuscitation phases (Li et al. 2016). A similar strategy was followed by Du et al. (2019), Fan et al. (2017), Yao et al. (2017, 2019), Wang et al. (2019d). (4) *Device-free*: There are not many popular datasets that are publicly available. However, researchers followed the same strategy which is adopted in RFID-based HAR. Yan et al. included the data of 6 volunteers with 440 actions in their dataset with a total of 4400 samples (Wang et al. 2019c). Similarly, Yan et al. (2020), Fei et al. (2020), Wang et al. (2019d) have proposed their own datasets.

## 6.4 Strengths and limitations

**Strengths** This is the first review of its kind where we demonstrated the HAR system consisting of three components such as HAR devices, AI models, and HAR applications. This is the only review that considered all four kinds of HAR devices such as sensor-based, vision-based, RFID-based, and device-free in the AI framework. The engineering perspective was discussed on AI in terms of architecture, loss function design, and optimization strategies. A comprehensive and comparative study was conducted in the benchmarking section. We also provided sources of datasets for the readers. *Limitations*: A significant amount of work has been done in the HAR domain, but some limitations need to be addressed. (1) *Synchronised activities*: According to earlier HAR models, researchers have made this presumption that a person performs a single activity at a time. But it is not true, in the real-world humans perform synchronized activities such as talking on smartphone and walking or reading a book. As per our knowledge, there is hardly a HAR model that considered synchronized activities in their recognition model. (2) *Complex and composite activities*: Various state-of-the-art results have been achieved by researchers with simple and atomic activities such as: running, walking, stairs-up, and down. But very limited work has been done with complex activities where an activity includes two or more simple actions. For example, exercise monitoring where an exercise like burpees includes jump, bending down, and extending legs. Such kind of complex and risky activity requires attention for proper posture monitoring, but to the best of our knowledge, there is no HAR model which can monitor exercise involving complex activity. (3) *Future action forecast*: A significant amount of work has been done in HAR but most of the work is based on the identification of action performed by a user like fall detection. There is no HAR model which predicts the future action. For example, in a smart home environment if an elderly person is doing exercise and there are chances of fall then it will be very helpful if there is a smart system that can identify fall in advance and inform the person timely for necessary precaution(s). (4) *Lack of real-time validation of HAR models*: In earlier HAR models, for

**Table 7** Previous reviews versus proposed review

C1	C2		C3				C4		C5				
	Citation	Year	HAR devices		AI type		Dataset (HAR-devices)						
			Sensor	Vision	RFID	DFP	ML	DL	Sensor	Vision	RFID	DFP	
R1	Hussain et al. (2020)	2020	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R2	Carvalho and Sofia (2020)	2020	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R3	Demrozi et al. (2020)	2020	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R4	Beddiar et al. (2020)	2020	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R5	Dhiman Chhavi (2019)	2019	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R6	Lima et al. (2019)	2019	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R7	De-La-Hoz-Franco (2018)	2018	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R8	Hx et al. (2017)	2017	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R9	Ke et al. (2013)	2013	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R10	Proposed	2021	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

DFP Device-free

validation researchers have used k-fold cross validation and LOSO, where a part on the dataset is used for validation. However, most of the data in datasets are gathered in the experimental setup, which lacks real-time flavour. Therefore, there is a need for a model which can provide good results on experimental as well as real-time data without AI bias (Suri et al. 2016).

## Recommendations

1. *Trending AI technique*: the use of transfer learning had shown significant results with *vision-based* HAR models. But there is very less work done on the *sensor-based* HAR model. *Sensor-based* HAR with TL can be the next revolution in the HAR domain.
2. *Trending device type*: a decade before the most popular data capturing source for activity signals were video cameras. But there are some major issues associated with *vision-based* HAR such as user privacy and GPU requirements. The solution to these problems is *sensor-based* HAR where a simple smartphone or smartwatch is used for capturing activity signals. In the last 3 years, the *sensor-based* HAR is one of the most trending HAR approach.
3. *Dominant target domain in HAR*: Although, HAR has multifaceted application domains such as surveillance, healthcare, and fall detection. Healthcare is the most crucial domain where HAR plays an important role which includes remote health monitoring to patients, exercise monitoring, assistive care to the elderly living alone. In the current COVID-19 pandemic scenario, the *sensor-based* HAR model with DL technique can be used to provide assistive care to home-quarantined COVID-19 patients by monitoring their health remotely.
4. *Abnormal action identification and future action prediction*: A significant amount of work has been done in HAR, but most of the work revolves around the recognition of simple activities. A very little amount of work has been done in finding the abnormalities in actions. Abnormality conditions are categorized into two categories: physical and non-physical. Under physical conditions, examples include (a) Fall detection in normal conditions under activities of daily living (ADL), (b) Fall detection in elderly health monitoring conditions, and (c) Fall detection in sports conditions. Only physical abnormality can be detected under this paradigm. Under Non-physical abnormality, examples include dizziness, headaches, vomiting feeling. These are not truly physical parameters that can be detected via the camera. Note that these non-physical parameters can however be monitored via special sensor-based devices, such as hypertension monitor, oximeter, etc. Further, to our knowledge, there are not many applications that combine camera and sensor devices in non-physical frame. Apart from abnormality identification, there is hardly any work done on the prediction of future action based on current actions. For example, A person is running or walking and he is not focusing or concentrating on the road on which he is travelling. Suddenly, there is an obstacle on the road in his path. He trips and falls down. Such detections are forecasting actions and happen suddenly. There is no application that can detect the obstacle and raise an alarm in advance. Forecasting is more towards the projections at distant times, unlike nearly current time spatial and temporal information. Similarly, forecasting is challenging in the motion estimation for subsequent frames where data is not available and unseen.

## 7 Conclusion

Unlike earlier review articles where researchers focus was on a single HAR device, we have proposed the study that revolves around the three pillars of HAR i.e., HAR devices, AI, and application domains. In the proposed review, we have hypothesized that the growth in HAR devices is synchronized with the evolving AI framework, and the study rationalizes this by providing evidence in terms of graphical representation of existing HAR models. Our second hypothesis says the growth in AI is the core of HAR which makes it suitable for multifaceted domains. We rationalized this by presenting representative CNN and TL architectures of HAR models, and also discussed the importance of hyperparameters, optimizers, and loss functions in the design of HAR models. A unique contribution is in the area of the role of the AI framework in existing HAR models for each of the HAR devices. This study further surfaced out (1) sensor-based HAR with miniaturizing devices will show the ground for opportunities in healthcare application, especially remote care, and monitoring, and (2) device-free HAR with the use of Wi-Fi device can make the usage of HAR as an essential part of human's healthy life. Finally, the study presented four recommendations that will expand the vision of new researchers and help them in expanding the scope of HAR in diverse domains with evolving AI framework for providing a quality of healthy life to human.

## Appendix A

The type of HAR devices and applications are two main components of HAR. Table [A.1](#), [A.2](#), [A.3](#), and [A.4](#) illustrates the device wise description of existing HAR models in terms of data source, #activities, #subjects, datasets, activity scenarios and performance evaluation.



**Table A.1** Sensor-based HAR

C1	C2	C3	C4	C5	C6	C7	C8
DS <sup>&amp;</sup>	Activities	Datasets	Subjects	Scenarios	Total # actions	Performance evaluation	CIT*
R1	WS ADL: 10, Sports: 11	Proprietary dataset	13- 10 M and 3 F	RT, Sports	1300/1100	99.65%, 99.92%	Hsu et al. (2018)
R2	SPS ADL: 6, 5	UCI-HAR, Weakly labelled (WL)	30 subs (UCI), 7 participants (WL)	Waist mounted SPS	76,157	UCI-HAR- 93.41%, WL- 93.83%	Wang et al. (2019a)
R3	SPS ADL: 7	Proprietary dataset	100 participants	Texting, handheld, trouser pocket, backpack	235 977 Samples	CNN-7-95.06%, Ensemble-96.11%	Zhu et al. (2019)
R4	SPS ADL: 4	Proprietary dataset	Active: 147, inactive: 99, walking: 200 and driving: 120. Total 574	Lab env	4,99,276	Mean accuracy-74.39%	Garcia-Gonzalez et al. (2020)
R5	SPS ADL: 6	HHAR	9 subs	Biking, walking, stairs	4,39,30,257	96.79%	Sundaramoorthy and Gudur (2018)
R6	SPS ADL: 6, 8, 9	HHAR, RWHAR, MobiAct	9, 15 (M-8, F-7), 57 (M-42, F-15)	Indoor and outdoor	4,39,30,257/2500	F1-measure on three datasets	Gouineua et al. (2018)
R7	SPS ADL: 8	RWHAR	15 subs	Experimental setup	8,85,360	F1-Score:0.94	Lawal and Bano (2020)
R8	SPS ADL: 6, 9, 14	HHAR, PAMAP, USC-HAD	9 (HHAR), 15 (RWHAR)	Waist mounted SPS, Experimental setup	4,39,30,257/38,50,505	F1-score: 0.848, 0.723, 0.702	Buffelli and Vandin (2020)
R9	SPS ADL: 6	HHAR	9 subs	Biking, walking, stairs	4,39,30,257	98%	Yao et al. (2019)
R10	SPS ADL: 6	HHAR, carTrack	9 subs	Biking, walking, stairs	4,39,30,257	94.2%	Yao et al. (2017)
R11	WS, SPS ADL: 6, 7, 17	UCI-HAR, WISDM, OPPORTUNITY	30 (UCI-HAR), 4 (OPPORTUNITY), 51 (WISDM)	Waist mounted SPS, Experimental setup, Biking, walking, stairs	76,157/2,551/15,630,426	F1-score: 92.63%, 95.85%, 95.78%	Xia et al. (2020)

Table A.1 (continued)

C1	C2	C3	C4	C5	C6	C7	C8
DS <sup>&amp;</sup>	Activities	Datasets	Subjects	Scenarios	Total # actions	Performance evaluation	CIT*
R12	SPS	ADL: 13, 5, 6	UniMibShar, MobiAct	Fall scenario	11,771/2500	87.30	Ferrari et al. (2020)
R13	SPS	ADL: 50	Vaizman dataset	Indoor, running (Phone in pocket)	300 k	92.80	Fazli et al. (2021)
R14	WS	19 (ADL)	13 subs (5 Female 8 male) of age (19–45 years)	9 activities indoor and 9 outdoor	–	Precision: 93.41%, recall: 93.16%	Pham et al. (2020)
R15	WS	ADL: (6, 12,6,11), GC	UCL-HAR, DaphNet, OPPORTUNITY, Skoda	ADL (UCL-HAR & Opportunity) Skoda: GC	76,157/15,630,426	96.7%, 97.8%, 92.5%, 94.1%, 92.6%	Murad and Pyun (2017)
R16	WS	18 ADL	13 subs of age (19–45 years)	9 indoor and 9 outdoor	–	F1-score: 77.7	Pham et al. (2017)

\*CIT citation

&amp; DS data source

WS wearable sensor, SPS smartphone sensor, Acc accelerometer, Gyro gyroscope, Mag magnetometer, ADL activities of daily living, RT Routine tasks, GC gestures in car maintenance

**Table A.2** Vision-based HAR

C1	C2	C3	C4	C5	C6	C7	C8
DS*	Activities	Datasets	Subjects	Scenarios	# Actions	Performance evaluation	CIT*
R1	DC 10, 10	UTKinect Action dataset, and CAD-60	10 (30 joints), 4	ADL: Home, kitchen, bedroom, bathroom, living room	60	97%, 96.5%	Phyo et al. (2019)
R2	DC 14, 20, 16	CAD-60, MSR Action3D and MSR Daily Activity 3D	4 (2 M-2F), 10, 10	Indoor	60/1000/567	94.12%, 86.81%, 68.75%	Qi et al. (2018)
R3	Vid 10	Weizmann	9	Non uniform background, walk	90	95.90%	Deep and Zheng (2019)
R4	Vid 101, 51	UCF-101 HMDB-51	YT, GV, MC	UCF-101: H-O, ADL, RT, HMDB-51: FE, H-O, H-H, ADL	13,320/7000	92.5%, 65.4%	Feichtenhofer et al. (2016)
R5	Vid 101, 51	UCF-101 HMDB-51	YT, GV, MC	UCF-101: H-O, ADL, RT, HMDB-51: FE, H-O, H-H, ADL	13,320/7000	UCF 101: 93.4%, HMDB 51: 66.4%	Feichtenhofer et al. (2017)
R6	Vid 400, 600	AVA, Kinetics 400, Kinetics-600, charades	MC and YT	H-O, H-H, ADL, RT	57,600/3,00,000/5,00,000	Kinetics: 75.6 and 92.1, Kinetics 600: 81.8, 95.1	Feichtenhofer et al. (2018)
R7	Vid -	AVA	Movie data	H-O, H-H, ADL, RT	57,600		Feichtenhofer and Ai (2019)
R8	Vid 101, 51, 400, 174	UCF-101, HMDB-51, kinetics-400, Something Something v1	YT, GV, MC	H-O, ADL, RT	13,320/7000/240 k/86,017	74.9%, 98.1%, 80.9%, 53.0%	Craato et al. (2019)

Table A.2 (continued)

C1	C2	C3	C4	C5	C6	C7	C8
DS <sup>&amp;</sup>	Activities	Datasets	Subjects	Scenarios	# Actions	Performance evaluation	CIT*
R9	Vid	Proposed HVU dataset (with YT-8 M), Kinetics-600 and HACS dataset	MC, YT	H-O, H-H	572 k videos—9 M annotation	–	Diba et al. (2020)
R10	Vid	UCF 101	YT	H-O, H-H, playing musical instruments, and sports	13,320	90.2%	Diba et al. (2016)
R11	Vid	UCSD: Ped1 and Ped 2, Avenue, Subway: entrance and exit	SVF	Crowd data from surveillance camera	–	90.5%, 88.9%, 90.3%, 91.6%, 98.4%	Wang et al. (2018)
R12	Vid	UCF-101, HMDB-51	YT, GV, MC	UCF-101: H-O, ADL, RT, HMDB-51: FE, H-O, H-H, ADL	13,320/7000	98%	Simonyan and Zisserman (2014)
R13	Vid	Something something v1 and v2, Jester, UCF-101, HMDB-51 and Kinetics-400	Actions while using objects	V1, v2: H-O, Jester: crowd acted,	108,499/220,887/148,092/300,000/13,320/6766	80.4, 89.8%, 99.9%, 91.6%, 96.2%, 72.2%	Jiang et al. (2019)
R14	Vid	UCSD Ped1, ped2	SVF	Crowd data of surveillance camera	70/28	97–98% accuracy on UMIN dataset	Thida et al. (2013)
R15	Vid	Weizmann, KTH, Ballet	ADL, Ballet dance	Weizmann: walk pattern from different angles, Ballet: DVD	90/6 × 100/44	–	Vishwakarma and Singh (2017)

Table A.2 (continued)

C1	C2	C3	C4	C5	C6	C7	C8
DS <sup>&amp;</sup>	Activities	Datasets	Subjects	Scenarios	# Actions	Performance evaluation	CIT*
R16	DC	17 (15 J), 4 Walk (20 T)	9 and 7 subjects	SPHERE: with three anomalies (short stop, stairs with right and stairs with left leg. DGD: four gait types	48/56	F1-score: SPHERE: 1, DGD: 0.98	Chaaraoui (2015)
R17	DC	FLIC, MPII	Elbow and wrist action images	FLIC: images from films, MPII: daily activity pose images	5003/25 k images	FLIC: 99% (elbow), 97% (wrist), MPII: 91.2% (elbow), 87.1% (wrist)	Newell Alejandro (2016)
R18	DC	20	Proprietary dataset	Daily activities	6220 frames	90.91%	Xia et al. (2012)

\*CIT citation, <sup>&</sup>DS data source, DC depth camera, Vid video, J joints, ADL activities of daily living, H-H Human to human interaction, H-O Human to object interaction, RT routine tasks, YT YouTube, MC Movie clip, GV Google video

Table A.3 RFID-based HAR

C1	C2	C3	C4	C5	C6	C7	C8
Data source	Activities	Datasets	Subjects	Scenarios	# Actions	Accuracy	CIT*
R1	RFID passive tags 10 free weight exercise	Proprietary dataset	15 (10 exercise)	Two weeks in a gym environment	Exercise Data 15 subs for 10 activities	90%	Ding et al. (2015)
R2	RFID passive tags 10 actions and 5 phases of resuscitation	Proprietary dataset	RSS during 16 actual trauma resuscitation	Realtime resuscitation	16 Trauma resuscitation data	80.2%	Li et al. (2016)
R3	RFID passive tags 23 orientation sensitive activities	Proprietary dataset	6 (5 F, 1 M)	Lab environment	23 actions performed by each subject for 120 secs	96%	Yao et al. (2018)
R4	RFID passive tags 10 ADL activities	Replaced sensors with RFID tags in Ordonez dataset	2 users' data of 21 days	Indoor environment, H-O	2747 instances	78.3%	Du et al. (2019)
R5	Smart Wall (passive rfid tags) 12 simple & complex ADL activities	Proprietary dataset	4	12 activities performed by 4 volunteers in mock room with SmartWall	–	97.9%	Oguntala et al. (2019)
R6	RFID tags 4 ADL	Proprietary dataset	10	Indoor	–	83.17%	Fan et al. (2017)

H-O human-object interaction, M male, F female, ADL activities of daily living, RSS received signal strength

\* CIT citation

**Table A.4** Device-free HAR

C1	C2	C3	C4	C5	C6	C7	C8
Data source	Activities	Datasets	Subs	Scenarios	# Actions	Accuracy	CIT*
R1 TL-WDR6500 AP Tx and Intel 5300 NIC card RP Rx	10 daily activities	Proprietary dataset	6 volunteers (3 male, 3 female)	Meeting room and lab room	4400	94.20%	Yan et al. (2020)
R2 Intel 5300 NIC	27 multi variation activities, Gaits: walk 10 m of volunteers	Proprietary dataset	5 volunteers (multi variation activities), 10 volunteers (of 10 walk gait data)	Lab environment	50 groups data for each volunteer	95%	Fei et al. (2020)
R3 Intel 5300 NIC	8 activities divided into 2 groups: torso based and gesture based	Proprietary dataset	Test model performance for 6 volunteers' data	Training in lab while testing in big hall, apartment and small office	5760	96%	Wang et al. (2019d)

NIC network interface card, Tx transmitter, Rx receiver, M male, F female

\*CIT citation

Table A.5 Sensor datasets

C1	C2	C3	C4	C5	C6	C7	C8			C9			
							Activity type				Single/ multiple person	Size	
Dataset name/ Publicly avail- able	Year	Source	# Classes	# Actors	Sensor location	FE	H-O	H-H	ADL	G			RT
R1	Skoda/Pub	2008	WS	10 gestures in car maintenance	1 subject	19 sensors on both arms	✓	✓	✓	✓	✓	Single	–
R2	USC-HAD/Pub	2012	IMU with Acc, Gyro, Mag	12 activities	17 (7 M, 7 F)	Front right hip	✓	✓	✓	✓	✓	Single	–
R3	PAMAP/Pub	2012	WS (IMU, 3 Colibri), HR monitor	18 activities	9 subjects	–	✓	✓	✓	✓	✓	Single	3,850,505 (52 attributes)
R4	Opportunity/Pub	2012	WS:(7 IMU, 12 Acc, 7 Loc), OS (12), AS (21)	6 runs per subject (5 ADL and 6th for drill)	4 subjects	Upper body, hip and leg	✓	✓	✓	✓	✓	Single	2551 (242 attributes)
R5	UCI-HAR/Pub	2012	SPS (Acc, Gyro)	6 activities	30 (19–48 years)	Samsung galaxy SII mounted on waist	✓	✓	✓	✓	✓	Single	10,299 (561 attributes)
R6	Heterogeneity (HHAR)/Pub	2015	SPS and SW Acc, Gyro	6 activities	9 users	8 SPS & 4 SW	✓	✓	✓	✓	✓	Single	43,930,257 (16 attributes)
R7	MobiAct/Pub	2016	SP (Acc, Gyro)	9 ADL activities and 4 types of falls	57 subjects (42 M, 15F) of (20–57 years)	Samsung Galaxy S3 SP in trousers' pocket	✓	✓	✓	✓	✓	Single	2500
R8	UniMibShar/Pub	2017	SPS	17 activities (9 ADL and 8 fall)	30 of (18–60 years)	–	✓	✓	✓	✓	✓	Single	11,771 samples
R9	WISDM/Pub	2019	SPS and SW's Acc, Gyro	18 activities	51	–	✓	✓	✓	✓	✓	Single	15,630,426



Table A.5 (continued)

C1	C2	C3	C4	C5	C6	C7	C8	C9		
Dataset name/ Publicly avail- able	Year	Source	# Classes	# Actors	Sensor location	Activity type	Single/ multiple person	Size		
						FE H-O H-H ADL G RT				
R10	19	NonSense/Pri	2020	e-Shoe, Sam- sung gear SW	18 activities (9 indoor and 9 outdoor sports)	13 (5F, 8 M) of (19–45 years)	Foot, arm	× × × × × × × × × × × × × × ×	Single	Not public dataset

\* CIT citations, SPS smartphone sensor, WS wearable sensor, SP smartphone, Acc accelerometer, Gyro gyroscope, Loc location, SW smartwatch, ADL activities of daily living, M male, F female, Pub publicly available, Prop proprietary

Table A.6 Video datasets

C1	C2	C3	C4	C5	C6	C7			C8	C9								
						Dataset name/ Publicly avail- able	Year	Source			# Classes	# Actor	Body part involved	Activity type			Single/multiple person	Size
														FE	H-O	H-H		
R1	CAD-60/Pub	2009	Kinect	5 (environ- ments) 12 (activities)	4 (2 M, 2F)	Whole body joint	✓	✓	✓	✓	✓	Single	60 videos					
R2	CAD-120/Pub	2009	Kinect	10 (High level) 10 (sub activ- ity labels) 12 (sub affor- dance labels)	4 (2 M, 2F)	Whole body joint	✓	✓	✓	✓	✓	Single	120 videos					
R3	MSR Action 3D/Pub	2009	DC	10 subjects, 20 action, 20 3D joints)	10	Whole body	✓	✓	✓	✓	✓	Single	336 action files					
R4	UT Kinect/Pub	2012	Kinect	10 actions	10	Whole body	✓	✓	✓	✓	✓	Single	1.79 GB					
R5	AVA/Pub	2018	MC	80 atomic visual actions	192 movies	Whole body	✓	✓	✓	✓	✓	Both	57,600 videos					
R6	UCF-101/Pub	2012	YT	101 actions	2,500 videos	Whole body	✓	✓	✓	✓	✓	Both	13 K clips 27 h					
R7	HMDB51/Pub	2011	YT, GV, MC	51 action classes	3,312 videos	Whole body	✓	✓	✓	✓	✓	Both	6,766 clips of 2 GB					
R8	Charades/Pub	2016	ADL	157 classes	267	Whole body	✓	✓	✓	✓	✓	Single	4855 KB					
R9	Kinetics 400/ Pub	2017	YT	400	400–1000 clips/ class	Whole body	✓	✓	✓	✓	✓	Both	3,00,000					
R10	Kinetics 600/ Pub	2018	YT	600	600–1000 clips/ class	Whole body	✓	✓	✓	✓	✓	Both	5,00,000					
R11	Something- Something/ Pub	2018	Objects Actions	174 classes	H-I actions	Whole body	✓	✓	✓	✓	✓	H-O interaction	108,499 videos					
R12	Weizmann/Pub	2005	ADL	10 action classes	2 subs	Whole body	✓	✓	✓	✓	✓	Single	90 videos					

**Table A.6** (continued)

C1	C2	C3	C4	C5	C6	C7	C8	C9
Dataset name/ Publicly avail- able	Year	Source	# Classes	# Actor	Body part involved	Activity type FE H-O H-H ADL G RT	Single/multiple person	Size
R13 UCSD/Pub	2013	camera	Peds1 and Peds2	Subway	People group	× × × × × ×	Surveillance data	Peds1: 60 & Peds2: 28

\* *CIT* citations, *ADL* activities of daily living, *M* male, *F* female, *YT* YouTube, *MC* movie clip, *DC* depth camera, *Pub* publicly available, *Prop* proprietary

## Appendix B

The performance of HAR model is evaluated using metrics. Table B.1 illustrates various evaluation metrics used in existing HAR models. But before the description of metrics, some terms need to be understood:

- True positive (TP): no. of positive samples predicted correctly.
- False positive (FP): no. of actual negative samples predicted as positive.
- True negative (TN): no. of negative samples predicted correctly.
- False negative (FN): no. of actual positive samples predicted as negative.

**Table B.1** Evaluation metrics

S. no.	Metrics	Description
1	Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$	Ratio of number of correct prediction and total number of input samples
2	Precision = $\frac{TP}{TP+FP}$	It is the no. of correct positives divided by the predicted positives
3	Recall = $\frac{TP}{TP+FN}$	It is the no. of correct positives divided by total no. of true positives and false negatives
4	F1 – score = $2 * \left[ \frac{P \times R}{P + R} \right]$	Harmonic mean between precision and recall
5	Specificity = $\frac{TN}{TN+FP}$	The proportion of actual negatives predicted as positives
6	Sensitivity = $\frac{TP}{TP+FN}$	The proportion of actual positives predicted as positives
7	Positive LHR = $\frac{\text{Sensitivity}}{100 - \text{Specificity}}$	LHR assess the goodness of fit of two competing statistical models based on their likelihoods
	Negative LHR = $\frac{100 - \text{Sensitivity}}{\text{Specificity}}$	

$P$  precision,  $R$  recall,  $TP$  true positive,  $TN$  true negative,  $FP$  false positive, and  $FN$  false negative,  $LHR$  likelihood ratio

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10462-021-10116-x>.

## References

- Abobakr A, Hossny M, Nahavandi S (2018) A skeleton-free fall detection system from depth images using random decision forest. *IEEE Syst J* 12(3):2994–3005. <https://doi.org/10.1109/JSYST.2017.2780260>
- Acharya UR et al (2012) An accurate and generalized approach to plaque characterization in 346 carotid ultrasound scans. *IEEE Trans Instrum Meas* 61(4):1045–1053. <https://doi.org/10.1109/TIM.2011.2174897>
- Acharya UR et al (2013b) Automated classification of patients with coronary artery disease using gray-scale features from left ventricle echocardiographic images. *Comput Methods Programs Biomed* 112(3):624–632. <https://doi.org/10.1016/j.cmpb.2013.07.012>
- Acharya UR et al (2015) Ovarian tissue characterization in ultrasound: a review. *Technol Cancer Res Treat* 14(3):251–261. <https://doi.org/10.1177/1533034614547445>
- Acharya UR, Sree SV, Saba L, Molinari F, Guerriero S, Suri JS (2013a) Ovarian tumor characterization and classification using ultrasound—a new online paradigm. *J Digit Imaging* 26(3):544–553. <https://doi.org/10.1007/s10278-012-9553-8>
- Adame T, Bel A, Carreras A, Melià-Seguí J, Oliver M, Pous R (2018) CUIDATS: An RFID–WSN hybrid monitoring system for smart health care environments. *Future Gen Comput Syst* 78:602–615. <https://doi.org/10.1016/j.future.2016.12.023>
- Agarwal M et al (2021) A novel block imaging technique using nine artificial intelligence models for COVID-19 disease classification, characterization and severity measurement in lung computed tomography scans on an Italian cohort. *J Med Syst*. <https://doi.org/10.1007/s10916-021-01707-w>
- Agarwal M et al (2021) Wilson disease tissue classification and characterization using seven artificial intelligence models embedded with 3D optimization paradigm on a weak training brain magnetic resonance imaging datasets: a supercomputer application. *Med Biol Eng Comput* 59(3):511–533. <https://doi.org/10.1007/s11517-021-02322-0>
- Akagündüz E, Aslan M, Şengür A (2016) Silhouette orientation volumes for efficient fall detection in depth videos. 2194(c):1–8. <https://doi.org/10.1109/JBHI.2016.2570300>
- Alsheikh MA, Selim A, Niyato D, Doyle L, Lin S, Tan HP (2016) Deep activity recognition models with triaxial accelerometers. In: AAAI workshop technical reports, vol. WS-16-01, pp 8–13, 2016.
- Arifoglu D, Bouchachia A (2017) Activity recognition and abnormal behaviour detection with recurrent neural networks. *Procedia Comput Sci* 110:86–93. <https://doi.org/10.1016/j.procs.2017.06.121>
- Asteriadis S, Daras P (2017) Landmark-based multimodal human action recognition. *Multimed Tools Appl* 76(3):4505–4521. <https://doi.org/10.1007/s11042-016-3945-6>
- Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y (2015) Physical human activity recognition using wearable sensors. *Sensors (Switzerland)* 15(12):31314–31338. <https://doi.org/10.3390/s151229858>
- Azkune G, Almeida A (2018) A scalable hybrid activity recognition approach for intelligent environments. *IEEE Access* 6(8):41745–41759. <https://doi.org/10.1109/ACCESS.2018.2861004>
- Bashar SK, Al Fahim A, Chon KH (2020) Smartphone based human activity recognition with feature selection and dense neural network. In: Proceedings of annual international conference of the ieee engineering in medicine and biology society EMBS, vol. 2020-July, pp 5888–5891, 2020. <https://doi.org/10.1109/EMBC44109.2020.9176239>
- 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>
- Biswas M et al (2018) Symtosis: a liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. *Comput Methods Programs Biomed* 155:165–177. <https://doi.org/10.1016/j.cmpb.2017.12.016>
- Biswas M et al (2019) State-of-the-art review on deep learning in medical imaging. *Front Biosci Landmark* 24(3):392–426. <https://doi.org/10.2741/4725>
- Buffelli D, Vandin F (2020) Attention-based deep learning framework for human activity recognition with user adaptation. *arXiv*, 2020.
- Cardoso HL, Mendes Moreira J (2016) Human activity recognition by means of online semi-supervised learning, pp. 75–77. <https://doi.org/10.1109/mdm.2016.93>
- Carreira J, Noland E, Banki-Horvath A, Hillier C, Zisserman A (2018) A short note about kinetics-600, 2018. [Online]. <http://arxiv.org/abs/1808.01340>.

- Carvalho LI, Sofia RC (2020) A review on scaling mobile sensing platforms for human activity recognition: challenges and recommendations for future research. *IoT* 1(2):451–473. <https://doi.org/10.3390/iot1020025>
- Chaaroui AA (2015) Abnormal gait detection with RGB-D devices using joint motion history features, 2015
- Chavarriga R et al (2013) The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. *Pattern Recognit Lett* 34(15):2033–2042. <https://doi.org/10.1016/j.patrec.2012.12.014>
- Chen WH, Cho PC, Jiang YL (2017) Activity recognition using transfer learning. *Sensors Mater* 29(7):897–904. <https://doi.org/10.18494/SAM.2017.1546>
- Chen Y, Shen C (2017) Performance analysis of smartphone-sensor behavior for human activity recognition. *IEEE Access* 5(c):3095–3110. <https://doi.org/10.1109/ACCESS.2017.2676168>
- Chen J, Sun Y, Sun S (2021) Improving human activity recognition performance by data fusion and feature engineering. *Sensors (Switzerland)* 21(3):1–23. <https://doi.org/10.3390/s21030692>
- Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y (2020) Deep learning for sensor-based human activity recognition: overview, challenges and opportunities. *arXiv*, vol. 37, no. 4, 2020
- Chong YS, Tay YH (2017) Abnormal event detection in videos using spatiotemporal autoencoder. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol 10262 LNCS, pp 189–196, 2017. [https://doi.org/10.1007/978-3-319-59081-3\\_23](https://doi.org/10.1007/978-3-319-59081-3_23)
- Cippitelli E, Gasparini S, Gambi E, Spinsante S (2016) A human activity recognition system using skeleton data from RGBD sensors. *Comput Intell Neurosci*. <https://doi.org/10.1155/2016/4351435>
- Civitaresse G, Presotto R, Bettini C (2019) Context-driven active and incremental activity recognition, 2019. [Online]. <http://arxiv.org/abs/1906.03033>.
- Cook DJ, Krishnan NC, Rashidi P (2013) Activity discovery and activity recognition: a new partnership. *IEEE Trans Cybern* 43(3):820–828. <https://doi.org/10.1109/TSMCB.2012.2216873>
- Cornell Activity Datasets: CAD-60 & CAD-120 (2021) [Online]. Available: [re3data.org](http://re3data.org): Cornell Activity Datasets: CAD-60 & CAD-120; editing status 2019-01-22; [re3data.org](http://re3data.org)—Registry of Research Data Repositories. <https://doi.org/10.17616/R3DD2D>. Accessed 17 Apr 2021
- Crasto N et al (2019) MARS: motion-augmented RGB stream for action recognition to cite this version : HAL Id : hal-02140558 MARS: motion-augmented RGB stream for action recognition, 2019. [Online]. <http://www.europe.naverlabs.com/Research/>
- Cui J, Xu B (2013) Cost-effective activity recognition on mobile devices. In: *BODYNETS 2013—8th international conference on body area networks*, pp 90–96, 2013. <https://doi.org/10.4108/icst.bodynets.2013.253656>
- De-La-Hoz-Franco E, Ariza-Colpas P, Quero JM, Espinilla M (2018) Sensor-based datasets for human activity recognition—a systematic review of literature. *IEEE Access* 6(c):59192–59210. <https://doi.org/10.1109/ACCESS.2018.2873502>
- Deep S, Zheng X (2019) Leveraging CNN and transfer learning for vision-based human activity recognition. In: 2019 29th international telecommunication networks and application conference ITNAC 2019, pp 35–38, 2019. <https://doi.org/10.1109/ITNAC46935.2019.9078016>
- Demrozi F, Pravadelli G, Bihorac A, Rashidi P (2020) Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. *IEEE Access* 8:210816–210836. <https://doi.org/10.1109/ACCESS.2020.3037715>
- Devanne M, Wannous H, Berretti S, Pala P, Daoudi M, Del Bimbo A (2015) 3-D human action recognition by shape analysis of motion trajectories on Riemannian manifold. *IEEE Trans Cybern* 45(7):1340–1352. <https://doi.org/10.1109/TCYB.2014.2350774>
- Dhiman Chhavi VDK (2019) state of art tech for HAR.pdf., pp 21–45
- Diba A, Pazandeh AM, Van Gool L (2016) Efficient two-stream motion and appearance 3D CNNs for video classification, 2016, [Online]. <http://arxiv.org/abs/1608.08851>
- Diba A et al. (2020) Large scale holistic video understanding. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol 12350 LNCS, pp 593–610, 2020. [https://doi.org/10.1007/978-3-030-58558-7\\_35](https://doi.org/10.1007/978-3-030-58558-7_35)
- Ding R et al (2019) Empirical study and improvement on deep transfer learning for human activity recognition. *Sensors (Switzerland)*. <https://doi.org/10.3390/s19010057>
- Ding W, Liu K, Fu X, Cheng F (2016) Profile HMMs for skeleton-based human action recognition. *Signal Process Image Commun* 42:109–119. <https://doi.org/10.1016/j.image.2016.01.010>
- Ding H et al. (2015) FEMO: a platform for free-weight exercise monitoring with RFIDs. In: *SenSys 2015—proceedings of 13th ACM conference on embedded networked sensor systems*, pp 141–154. <https://doi.org/10.1145/2809695.2809708>.

- Du Y, Lim Y, Tan Y (2019) A novel human activity recognition and prediction in smart home based on interaction. *Sensors* (Switzerland). <https://doi.org/10.3390/s19204474>
- Duan H, Zhao Y, Xiong Y, Liu W, Lin D (2020) Omni-sourced webly-supervised learning for video recognition. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol 12360 LNCS, pp 670–688, 2020. [https://doi.org/10.1007/978-3-030-58555-6\\_40](https://doi.org/10.1007/978-3-030-58555-6_40)
- Ehatisham-Ul-Haq M, Azam MA, Amin Y, Naeem U (2020) C2FHAR: coarse-to-fine human activity recognition with behavioral context modeling using smart inertial sensors. *IEEE Access* 8:7731–7747. <https://doi.org/10.1109/ACCESS.2020.2964237>
- El-Baz JSSA, Jiang X (2016) *Biomedical Image Segmentation: Advances and Trends*. CRC Press, Taylor & Francis Group
- El-Baz A, Suri JS (2019) *Level set method in medical imaging segmentation*. CRC Press, Taylor & Francis Group, London
- Fan X, Gong W, Liu J (2017) I2tag: RFID mobility and activity identification through intelligent profiling. *ACM Trans Intell Syst Technol* 9(1):1–21. <https://doi.org/10.1145/3035968>
- Fan X, Wang F, Wang F, Gong W, Liu J (2019) When RFID meets deep learning: exploring cognitive intelligence for activity identification. *IEEE Wirel Commun* 26(3):19–25. <https://doi.org/10.1109/MWC.2019.1800405>
- Fazli M, Kowsari K, Gharavi E, Barnes L, Doryab A (2020) HHAR-net: hierarchical human activity recognition using neural networks, pp 48–58, 2021. [https://doi.org/10.1007/978-3-030-68449-5\\_6](https://doi.org/10.1007/978-3-030-68449-5_6)
- Fei H, Xiao F, Han J, Huang H, Sun L (2020) Multi-variations activity based gaits recognition using commodity WiFi. *IEEE Trans Veh Technol* 69(2):2263–2273. <https://doi.org/10.1109/TVT.2019.2962803>
- Feichtenhofer C, Ai F (2019) SlowFast networks for video recognition technical report AVA action detection in ActivityNet challenge 2019, pp. 2–5
- Feichtenhofer C, Fan H, Malik J, He K (2018) SlowFast networks for video recognition 2018. [Online]. <http://arxiv.org/abs/1812.03982>
- Feichtenhofer C, Pinz A, Wildes RP (2017) Spatiotemporal multiplier networks for video action recognition. In: *Proceedings of 30th IEEE conference on computer vision and pattern recognition, CVPR 2017*, vol 2017-Janua, no. Nips, pp 7445–7454, 2017. <https://doi.org/10.1109/CVPR.2017.787>
- Feichtenhofer C, Pinz A, Zisserman A (2016) Convolutional two-stream network fusion for video action recognition. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, vol 2016-Decem, no. i, pp. 1933–1941, 2016. <https://doi.org/10.1109/CVPR.2016.213>
- Ferrari A, Micucci D, Mobilio M, Napoletano P (2020) On the personalization of classification models for human activity recognition. *IEEE Access* 8:32066–32079. <https://doi.org/10.1109/ACCESS.2020.2973425>
- Fullerton E, Heller B, Munoz-Organero M (2017) Recognizing human activity in free-living using multiple body-worn accelerometers. *IEEE Sens J* 17(16):5290–5297. <https://doi.org/10.1109/JSEN.2017.2722105>
- Gaglio S, Lo Re G, Morana M (2015) Human activity recognition process using 3-D posture data. *IEEE Trans Hum Mach Syst* 45(5):586–597. <https://doi.org/10.1109/THMS.2014.2377111>
- Gani MO et al (2019) A light weight smartphone based human activity recognition system with high accuracy. *J Netw Comput Appl* 141(May):59–72. <https://doi.org/10.1016/j.jnca.2019.05.001>
- Garcia-Gonzalez D, Rivero D, Fernandez-Blanco E, Luaces MR (2020) A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors* (Switzerland). <https://doi.org/10.3390/s20082200>
- Gorelick L, Blank M, Shechtman E, Irani M, Basri R (2007) Actions as space-time shapes. *IEEE Trans Pattern Anal Mach Intell* 29(12):2247–2253. <https://doi.org/10.1109/TPAMI.2007.70711>
- Gouineua F, Sortin M, Chikhaoui B (2018) Chikhaoui-DL-springer (2018).pdf. Springer, pp 302–315
- Goyal R et al. (2017) The ‘Something Something’ video database for learning and evaluating visual common sense. In: *Proceedings of the IEEE international conference on computer vision*, pp 5843–5851. <https://doi.org/10.1109/ICCV.2017.622>
- Gu C et al. (2018) AVA: a video dataset of spatio-temporally localized atomic visual actions. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, pp 6047–6056, 2018. <https://doi.org/10.1109/CVPR.2018.00633>
- Han J et al. (2014) CBID: a customer behavior identification system using passive tags. In: *Proceedings of international conference on network protocols, ICNP*, pp 47–58, 2014. <https://doi.org/10.1109/ICNP.2014.26>

- Hsu YL, Yang SC, Chang HC, Lai HC (2018) Human daily and sport activity recognition using a wearable inertial sensor network. *IEEE Access* 6(c):31715–31728. <https://doi.org/10.1109/ACCESS.2018.2839766>
- Huang SF, Chang RF, Moon WK, Lee YH, Chen DR, Suri JS (2008) Analysis of tumor vascularity using ultrasound images. *IEEE Trans Med Imaging* 27(3):320–330
- Hussain Z, Sheng QZ, Zhang WE (2020) A review and categorization of techniques on device-free human activity recognition. *J Netw Comput Appl* 167:102738. <https://doi.org/10.1016/j.jnca.2020.102738>
- Hx P, Wang J, Hu L, Chen Y, Hao S (2017) Deep learning for sensor based activity recognition: a survey. *Pattern Recognit Lett* 1–9
- Jalal A, Uddin M, Kim TS (2012) Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE Trans Consum Electron* 58(3):863–871. <https://doi.org/10.1109/TCE.2012.6311329>
- Jamthikar AD et al (2020) Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study. *Int J Cardiovasc Imaging*. <https://doi.org/10.1007/s10554-020-02099-7>
- Janocha K, Czarniecki WM (2016) On loss functions for deep neural networks in classification. *Schedae Informaticae* 25:49–59. <https://doi.org/10.4467/20838476SI.16.004.6185>
- Jiang B, Wang M, Gan W, Wu W, Yan J (2019) STM: spatiotemporal and motion encoding for action recognition. In: *Proceedings of the IEEE international conference on computer vision*, vol. 2019–Octob, pp 2000–2009, 2019. <https://doi.org/10.1109/ICCV.2019.00209>
- Kalfaoglu ME, Kalkan S, Alatan AA (2020) Late temporal modeling in 3D CNN architectures with bert for action recognition. *arXiv*, pp 1–19. [https://doi.org/10.1007/978-3-030-68238-5\\_48](https://doi.org/10.1007/978-3-030-68238-5_48)
- Kasnesis P, Patrikakis CZ, Venieris IS (2017) Changing mobile data analysis through deep learning, pp 17–23
- Kay W et al. (2017) The kinetics human action video dataset, 2017 [Online]. <http://arxiv.org/abs/1705.06950>
- Ke SR, Thuc HLU, Lee YJ, Hwang JN, Yoo JH, Choi KH (2013) A review on video-based human activity recognition, vol 2, no 2
- Khalifa S, Lan G, Hassan M, Seneviratne A, Das SK (2018) HARKE: human activity recognition from kinetic energy harvesting data in wearable devices. *IEEE Trans Mob Comput* 17(6):1353–1368. <https://doi.org/10.1109/TMC.2017.2761744>
- Kim S, Yun K, Park J, Choi JY (2019) Skeleton-based action recognition of people handling objects. In: *Proceedings of 2019 IEEE winter conference on applications of computer vision, WACV 2019*, pp 61–70, 2019. <https://doi.org/10.1109/WACV.2019.00014>
- Koppula HS, Gupta R, Saxena A (2013) Learning human activities and object affordances from RGB-D videos. *Int J Rob Res* 32(8):951–970. <https://doi.org/10.1177/0278364913478446>
- Koppula HS, Saxena A (2016) Anticipating human activities using object affordances for reactive robotic response. *IEEE Trans Pattern Anal Mach Intell* 38(1):14–29. <https://doi.org/10.1109/TPAMI.2015.2430335>
- Kuehne H, Jhuang H, Garrote E, Poggio T, Serre T (2011) HMDB: a large video database for human motion recognition. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2556–2563. <https://doi.org/10.1109/ICCV.2011.6126543>
- Lara ÓD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor* 15(3):1192–1209. <https://doi.org/10.1109/SURV.2012.110112.00192>
- Lawal IA, Bano S (2020) Deep human activity recognition with localisation of wearable sensors. *IEEE Access* 8:155060–155070. <https://doi.org/10.1109/ACCESS.2020.3017681>
- Lawal IA, Bano S (2019) Deep human activity recognition using wearable sensors. In: *ACM international conference proceedings series*, pp 45–48, 2019. <https://doi.org/10.1145/3316782.3321538>
- Li JH, Tian L, Wang H, An Y, Wang K, Yu L (2019) Segmentation and recognition of basic and transitional activities for continuous physical human activity. *IEEE Access* 7:42565–42576. <https://doi.org/10.1109/ACCESS.2019.2905575>
- Li W, Zhang Z, Liu Z (2010) Action recognition based on a bag of 3D points. In: *2010 IEEE computer society conference on computer vision and pattern recognition—work. CVPRW 2010*, vol 2010, pp 9–14, 2010. <https://doi.org/10.1109/CVPRW.2010.5543273>
- Li X, Zhang Y, Marsic I, Sarcevic A, Burd RS (2016) Deep learning for RFID-based activity recognition. In: *Proceedings of 14th ACM conference on embedded networked sensor systems SenSys 2016*, pp 164–175. <https://doi.org/10.1145/2994551.2994569>



- Lima WS, Souto E, El-Khatib K, Jalali R, Gama J (2019) Human activity recognition using inertial sensors in a smartphone: an overview. *Sensors (switzerland)* 19(14):14–16. <https://doi.org/10.3390/s19143213>
- Liu J, Shahroudy A, Xu D, Wang G (2016) Spatio-temporal LSTM with trust gates for 3D human action recognition. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol 9907 LNCS, pp 816–833, 2016. [https://doi.org/10.1007/978-3-319-46487-9\\_50](https://doi.org/10.1007/978-3-319-46487-9_50).
- Liu Z, Zhang H, Chen Z, Wang Z, Ouyang W (2020) Disentangling and unifying graph convolutions for skeleton-based action recognition. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, pp 140–149, 2020. <https://doi.org/10.1109/CVPR42600.2020.00022>
- Lv T, Wang X, Jin L, Xiao Y, Song M (2020) A hybrid network based on dense connection and weighted feature aggregation for human activity recognition. *IEEE Access* 8:68320–68332. <https://doi.org/10.1109/ACCESS.2020.2986246>
- Mabrouk MF, Ghanem NM, Ismail MA (2016) Semi supervised learning for human activity recognition using depth cameras. In: *Proceedings of 2015 IEEE 14th international conference on machine learning and applications ICMLA 2015*, pp 681–686, 2016. <https://doi.org/10.1109/ICMLA.2015.170>
- Magherini T, Fantechi A, Nugent CD, Vicario E (2013) Using temporal logic and model checking in automated recognition of human activities for ambient-assisted living. *IEEE Trans Hum Mach Syst* 43(6):509–521. <https://doi.org/10.1109/TSMC.2013.2283661>
- Mahadevan V, Li W, Bhalodia V, Vasconcelos N (2010) Anomaly detection in crowded scenes. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, pp 1975–1981, 2010. <https://doi.org/10.1109/CVPR.2010.5539872>
- Maniruzzaman M et al (2017) Comparative approaches for classification of diabetes mellitus data: machine learning paradigm. *Comput Methods Programs Biomed* 152:23–34. <https://doi.org/10.1016/j.cmpb.2017.09.004>
- Maniruzzaman M et al (2018) Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *J Med Syst* 42(5):1–17. <https://doi.org/10.1007/s10916-018-0940-7>
- Martis JSRJ, Acharya UR, Prasad H, Chua CK, Lim CM (2013) Application of higher order statistics for atrial arrhythmia classification. *Biomed Signal Process Control* 8(6)
- Micucci D, Mobilio M, Napoletano P (2017) UniMiB SHAR: a dataset for human activity recognition using acceleration data from smartphones. *Appl Sci*. <https://doi.org/10.3390/app7101101>
- Miu T, Missier P, Plötz T (2015) Bootstrapping personalised human activity recognition models using online active learning. In: *Proceedings of 15th international conference on computer science and information technology CIT 2015, 14th IEEE international conference on ubiquitous computing and communications IUCC 2015, 13th international conference on dependable, autonomic and secure*, pp 1138–1147, 2015. <https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.170>
- Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies (2011)
- Munoz-Organero M (2019) Outlier detection in wearable sensor data for human activity recognition (HAR) based on DRNNs. *IEEE Access* 7:74422–74436. <https://doi.org/10.1109/ACCESS.2019.2921096>
- Murad A, Pyun JY (2017) Deep recurrent neural networks for human activity recognition. *Sensors (Switzerland)*. <https://doi.org/10.3390/s17112556>
- Nam Y, Park JW (2013) Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor. *IEEE J Biomed Heal Inform* 17(2):420–426. <https://doi.org/10.1109/JBHI.2012.2235075>
- Nash W, Drummond T, Birbilis N (2018) A review of deep learning in the study of materials degradation. *NPJ Mater Degrad* 2(1):1–12. <https://doi.org/10.1038/s41529-018-0058-x>
- Neili Boualia S, Essoukri Ben Amara N (2021) Deep full-body HPE for activity recognition from RGB frames only. *Informatics* 8(1):2. <https://doi.org/10.3390/informatics8010002>
- Newell Alejandro DJ, Yang K (2016) Stacked hour glass.pdf., pp 1–15
- Nguyen DT, Kim KW, Hong HG, Koo JH, Kim MC, Park KR (2017) Gender recognition from human-body images using visible-light and thermal camera videos based on a convolutional neural network for image feature extraction, pp 1–22, 2017. <https://doi.org/10.3390/s17030637>
- Ni B, Pei Y, Moulin P, Yan S (2013) Multilevel depth and image fusion for human activity detection. *IEEE Trans Cybern* 43(5):1382–1394. <https://doi.org/10.1109/TCYB.2013.2276433>
- Obaida MA, Saraee MAM (2017) A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. *Neural Comput Appl* 28(s1):565–572. <https://doi.org/10.1007/s00521-016-2363-z>


- Oguntala GA et al (2019) SmartWall: novel RFID-enabled ambient human activity recognition using machine learning for unobtrusive health monitoring. *IEEE Access* 7:68022–68033. <https://doi.org/10.1109/ACCESS.2019.2917125>
- Ohn-Bar E, Trivedi MM (2014) Hand gesture recognition in real time for automotive interfaces: a multimodal vision-based approach and evaluations. *IEEE Trans Intell Transp Syst* 15(6):2368–2377. <https://doi.org/10.1109/TITS.2014.2337331>
- Ordóñez FJ, Roggen D (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors (Switzerland)*. <https://doi.org/10.3390/s16010115>
- Parada R, Nur K, Melia-Segui J, Pous R (2016) Smart surface: RFID-based gesture recognition using k-means algorithm. In: *Proceedings of 12th international conference on intelligent environments IE 2016*, pp 111–118, 2016. <https://doi.org/10.1109/IE.2016.25>
- Pareek G et al (2013) Prostate tissue characterization/classification in 144 patient population using wavelet and higher order spectra features from transrectal ultrasound images. *Technol Cancer Res Treat* 12(6):545–557. <https://doi.org/10.7785/tcrt.2012.500346>
- Pham C et al (2020) SensCapsNet: deep neural network for non-obtrusive sensing based human activity recognition. *IEEE Access* 8:86934–86946. <https://doi.org/10.1109/ACCESS.2020.2991731>
- Pham C, Diep NN, Phuong TM (2017) E-shoes: smart shoes for unobtrusive human activity recognition. In: *Proceedings of 2017 9th international conference on knowledge and systems engineering KSE 2017*, vol 2017-Janua, pp 269–274, 2017. <https://doi.org/10.1109/KSE.2017.8119470>
- Phyo CN, Zin TT, Tin P (2019) Deep learning for recognizing human activities using motions of skeletal joints. *IEEE Trans Consum Electron* 65(2):243–252. <https://doi.org/10.1109/TCE.2019.2908986>
- Popoola OP, Wang K (2012) Video-based abnormal human behavior recognition review. *IEEE Trans Syst Man Cybern Part C Appl Rev* 42(6):865–878. <https://doi.org/10.1109/TSMCC.2011.2178594>
- Qi J, Wang Z, Lin X, Li C (2018) Learning complex spatio-temporal configurations of body joints for online activity recognition. *IEEE Trans Hum Mach Syst* 48(6):637–647. <https://doi.org/10.1109/THMS.2018.2850301>
- Qin Z, Zhang Y, Meng S, Qin Z, Choo KKR (2020) Imaging and fusing time series for wearable sensor-based human activity recognition. *Inf Fusion* 53:80–87. <https://doi.org/10.1016/j.inffus.2019.06.014>
- Raad MW, Sheltami T, Soliman MA, Alrashed M (2018) An RFID based activity of daily living for elderly with Alzheimer's. In: *Lecture notes of the institute for computer sciences, social-informatics and telecommunications engineering LNICST*, vol 225, pp 54–61, 2018. [https://doi.org/10.1007/978-3-319-76213-5\\_8](https://doi.org/10.1007/978-3-319-76213-5_8)
- Rajendra Acharya U et al (2014) A review on ultrasound-based thyroid cancer tissue characterization and automated classification. *Technol Cancer Res Treat* 13(4):289–301. <https://doi.org/10.7785/tcrt.2012.500381>
- Rashidi P, Mihailidis A (2013) A survey on ambient-assisted living tools for older adults. *IEEE J Biomed Heal Inform* 17(3):579–590. <https://doi.org/10.1109/JBHI.2012.2234129>
- Ravi D, Wong C, Lo B, Yang GZ (2016) Deep learning for human activity recognition: a resource efficient implementation on low-power devices. In: *BSN 2016—13th annual body sensor networks conference*, pp 71–76, 2016. <https://doi.org/10.1109/BSN.2016.7516235>
- Reiss A, Stricker D (2012) Introducing a new benchmarked dataset for activity monitoring. In: *Proceedings of international symposium on wearable computers ISWC*, pp 108–109, 2012. <https://doi.org/10.1109/ISWC.2012.13>
- Reiss A, Stricker D (2012) Creating and benchmarking a new dataset for physical activity monitoring. In: *ACM international conference proceeding series*, no. February, 2012. <https://doi.org/10.1145/2413097.2413148>
- Roggen D et al (2010) “Collecting complex activity datasets in highly rich networked sensor environments”, *INSS 2010—7th Int. Conf Networked Sens Syst* 00:233–240. <https://doi.org/10.1109/INSS.2010.5573462>
- Ronao CA, Cho SB (2016) Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst Appl* 59:235–244. <https://doi.org/10.1016/j.eswa.2016.04.032>
- Röcker C, O'Donoghue J, Zieffle M, Maciaszek L, Molloy W (2017) Preface. *Commun Comput. Inf Sci* 736:5. <https://doi.org/10.1007/978-3-319-62704-5>
- Saba L et al (2019) The present and future of deep learning in radiology. *Eur J Radiol* 114:14–24. <https://doi.org/10.1016/j.ejrad.2019.02.038>
- Saba L et al (2021) Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system. *Int J Cardiovasc Imaging*. <https://doi.org/10.1007/s10554-020-02124-9>
- Saha J, Ghosh D, Chowdhury C, Bandyopadhyay S (2020) Smart handheld based human activity recognition using multiple instance multiple label learning. *Wirel Pers Commun*. <https://doi.org/10.1007/s11277-020-07903-0>

- Shivendra shivani JSS, Agarwal S (2018) Hand book of image-based security techniques. Chapman and Hall/CRC, London, p 442
- Shrivastava VK, Londhe ND, Sonawane RS, Suri JS (2016) Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: a first comparative study of its kind. *Comput Methods Programs Biomed* 126(2016):98–109. <https://doi.org/10.1016/j.cmpb.2015.11.013>
- Shuaibu AN, Malik AS, Faye I, Ali YS (2017) Pedestrian group attributes detection in crowded scenes. In: Proceedings of 3rd international conference on advanced technologies for signal and image processing ATSIP 2017, pp 1–5, 2017. <https://doi.org/10.1109/ATSIP.2017.8075584>
- Sigurdsson GA, Varol G, Wang X, Farhadi A, Laptev I, Gupta A (2016) Hollywood in homes: crowdsourcing data collection for activity understanding. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 9905 LNCS, pp 510–526, 2016. [https://doi.org/10.1007/978-3-319-46448-0\\_31](https://doi.org/10.1007/978-3-319-46448-0_31).
- Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. *Adv Neural Inf Process Syst* 1:568–576
- Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 3rd international conference on learning representations ICLR 2015—conference track proceedings, pp 1–14
- Skandha SS et al (2020) 3-D optimized classification and characterization artificial intelligence paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: Atheromatic™ 2.0. *Comput Biol Med* 125:103958. <https://doi.org/10.1016/j.combiomed.2020.103958>
- Soomro K, Zamir AR, Shah M (2012) UCF101: a dataset of 101 human actions classes from videos in the wild, no. November, 2012, [Online]. <http://arxiv.org/abs/1212.0402>.
- Soydaner D (2020) A comparison of optimization algorithms for deep learning. *Int J Pattern Recognit Artif Intell*. <https://doi.org/10.1142/S0218001420520138>
- Sreekanth NS, Narayanan NK (2017) Proceedings of the international conference on signal, networks, computing, and systems, vol 395, pp 105–115, 2017. <https://doi.org/10.1007/978-81-322-3592-7>
- Stisen A et al. (2015) Smart devices are different: assessing and mitigating mobile sensing heterogeneities for activity recognition. In: SenSys 2015—proceedings of 13th ACM conference on embedded networked sensor systems, no. November, pp 127–140, 2015. <https://doi.org/10.1145/2809695.2809718>
- Sudeep PV et al (2016) Speckle reduction in medical ultrasound images using an unbiased non-local means method. *Biomed Signal Process Control* 28:1–8. <https://doi.org/10.1016/j.bspc.2016.03.001>
- Sun S, Cao Z, Zhu H, Zhao J (2020) A survey of optimization methods from a machine learning perspective. *IEEE Trans Cybern* 50(8):3668–3681. <https://doi.org/10.1109/TCYB.2019.2950779>
- Sundaramoorthy P, Gudur GK (2018) HARNet : towards on-device incremental learning using deep, pp 31–36
- Sung J, Ponce C, Selman B, Saxena A (2012) Unstructured human activity detection from RGBD images. In: Proceedings of IEEE international conference on robotics and automation, pp 842–849, 2012. <https://doi.org/10.1109/ICRA.2012.6224591>
- Suri JS (2001) Two-dimensional fast magnetic resonance brain segmentation. *IEEE Eng Med Biol Mag* 20(4):84–95. <https://doi.org/10.1109/51.940054>
- Suri JS (2005) Handbook of biomedical image analysis: segmentation models. Springer, New York
- Suri JS et al (2021) Systematic review of artificial intelligence in acute respiratory distress syndrome for COVID-19 lung patients: a biomedical imaging perspective. *IEEE J Biomed Heal Inform* 2194(1):1–12. <https://doi.org/10.1109/JBHI.2021.3103839>
- Suri JS, Liu K, Singh S, Laxminarayan SN, Zeng X, Reden L (2002) Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review. *IEEE Trans Inf Technol Biomed* 6(1):8–28. <https://doi.org/10.1109/4233.992158>
- Suri JS (2013) DK Med\_Image\_Press\_Eng.Pdf.” [Online]. <https://www.freepatentsonline.com/20080051648.pdf>.
- Suri JS (2004) Segmentation method and apparatus for medical images using diffusion propagation, pixel classification, and mathematical morphology
- Suthar B, Gadhia B (2021) Human activity recognition using deep learning: a survey. *Lect Notes Data Eng Commun Technol* 52:217–223. [https://doi.org/10.1007/978-981-15-4474-3\\_25](https://doi.org/10.1007/978-981-15-4474-3_25)
- Szegedy C et al (2015) Going deeper with convolutions. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 2818–2826, 2016. <https://doi.org/10.1109/CVPR.2016.308>.

- Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI conference on artificial intelligence AAAI 2017, pp. 4278–4284
- Tanberk S, Kilimci ZH, Tukul DB, Uysal M, Akyokus S (2020) A hybrid deep model using deep learning and dense optical flow approaches for human activity recognition. *IEEE Access* 8:19799–19809. <https://doi.org/10.1109/ACCESS.2020.2968529>
- Tandel GS, Balestrieri A, Jujaray T, Khanna NN, Saba L, Suri JS (2020) Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. *Comput Biol Med* 122:103804. <https://doi.org/10.1016/j.compbiomed.2020.103804>
- Tao D, Jin L, Yuan Y, Xue Y (2016a) Ensemble manifold rank preserving for acceleration-based human activity recognition. *IEEE Trans Neural Networks Learn Syst* 27(6):1392–1404. <https://doi.org/10.1109/TNNLS.2014.2357794>
- Tao D, Wen Y, Hong R (2016b) Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition. *IEEE Internet Things J* 3(6):1124–1134. <https://doi.org/10.1109/JIOT.2016.2561962>
- Thida M, Eng HL, Remagnino P (2013) Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes. *IEEE Trans Cybern* 43(6):2147–2156. <https://doi.org/10.1109/TCYB.2013.2242059>
- Tian Y, Zhang J, Chen L, Geng Y, Wang X (2019) Single wearable accelerometer-based human activity recognition via kernel discriminant analysis and QPSO-KELM classifier. *IEEE Access* 7:109216–109227. <https://doi.org/10.1109/access.2019.2933852>
- Tran D, Wang H, Feiszli M, Torresani L (2019) Video classification with channel-separated convolutional networks. In: Proceedings of IEEE international conference on computer vision, vol 2019-October, pp 5551–5560, 2019. <https://doi.org/10.1109/ICCV.2019.00565>.
- Vaniya SM, Bharathi B (2017) Exploring object segmentation methods in visual surveillance for human activity recognition. In: Proceedings of International Conference on Global Trends in Signal Processing, Information Computing and Communication. ICGTSPICC 2016, pp 520–525, 2017. <https://doi.org/10.1109/ICGTSPICC.2016.7955356>
- Vishwakarma DK, Singh K (2017) Human activity recognition based on spatial distribution of gradients at sublevels of average energy silhouette images. *IEEE Trans Cogn Dev Syst* 9(4):316–327. <https://doi.org/10.1109/TCDS.2016.2577044>
- Wang A, Chen G, Yang J, Zhao S, Chang CY (2016a) A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J* 16(11):4566–4578. <https://doi.org/10.1109/JSEN.2016.2545708>
- Wang F, Feng J, Zhao Y, Zhang X, Zhang S, Han J (2019c) Joint activity recognition and indoor localization with WiFi fingerprints. *IEEE Access* 7:80058–80068. <https://doi.org/10.1109/ACCESS.2019.2923743>
- Wang F, Gong W, Liu J (2019d) On spatial diversity in wifi-based human activity recognition: a deep learning-based approach. *IEEE Internet Things J* 6(2):2035–2047. <https://doi.org/10.1109/JIOT.2018.2871445>
- Wang K, He J, Zhang L (2019a) Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors. *IEEE Sens J* 19(17):7598–7604. <https://doi.org/10.1109/JSEN.2019.2917225>
- Wang Q, Ma Y, Zhao K, Tian Y (2020) A comprehensive survey of loss functions in machine learning. *Ann Data Sci*. <https://doi.org/10.1007/s40745-020-00253-5>
- Wang Z, Wu D, Chen J, Ghoneim A, Hossain MA (2016b) A triaxial accelerometer-based human activity recognition via EEMD-based features and game-theory-based feature selection. *IEEE Sens J* 16(9):3198–3207. <https://doi.org/10.1109/JSEN.2016.2519679>
- Wang F, Liu J, Gong W (2020) Multi-adversarial in-car activity recognition using RFIDs. *IEEE Trans Mob Comput* 1–1. <https://doi.org/10.1109/tmc.2020.2977902>
- Wang X, Ji Q (2014) A hierarchical context model for event recognition in surveillance video. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 2561–2568. <https://doi.org/10.1109/CVPR.2014.328>.
- Wang K, He J, Zhang L (2019) Attention-based convolutional neural network for weakly labeled human activities recognition with wearable sensors. *arXiv*, vol 19, no. 17, pp 7598–7604
- Wang L, Zhou F, Li Z, Zuo W, Tan H (2018) Abnormal event detection in videos using hybrid spatio-temporal autoencoder school of instrumentation science and opto-electronics Engineering, Beihang University, Beijing, China Department of Electronic Information Engineering, Foshan University, Fo. In: 2018 25th IEEE international conference on image processing, pp 2276–2280

- Weiss GM, Yoneda K, Hayajneh T (2019) Smartphone and smartwatch-based biometrics using activities of daily living. *IEEE Access* 7:133190–133202. <https://doi.org/10.1109/ACCESS.2019.2940729>
- Weng Z, Li W, Jin Z (2021) Human activity prediction using saliency-aware motion enhancement and weighted LSTM network. *Eurasip J Image Video Process* 1:2021. <https://doi.org/10.1186/s13640-020-00544-0>
- Xia K, Huang J, Wang H (2020) LSTM-CNN architecture for human activity recognition. *IEEE Access* 8:56855–56866. <https://doi.org/10.1109/ACCESS.2020.2982225>
- Xia L, Chen C, Aggarwal J (2012) View invariant human action recognition using histograms of 3D joints The University of Texas at Austin. In: *CVPR 2012 HAU3D workshop*, pp 20–27, 2012, [Online]. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:View+Invariant+Human+Action+Recognition+Using+Histograms+of+3D+Joints+The+University+of+Texas+at+Austin#1>
- Xie L, Wang C, Liu AX, Sun J, Lu S (2018) Multi-Touch in the air: concurrent micromovement recognition using RF signals. *IEEE/ACM Trans Netw* 26(1):231–244. <https://doi.org/10.1109/TNET.2017.2772781>
- Xu W, Miao Z, Zhang XP, Tian Y (2017) A hierarchical spatio-temporal model for human activity recognition. *IEEE Trans Multimed* 19(7):1494–1509. <https://doi.org/10.1109/TMM.2017.2674622>
- Xu X, Tang J, Zhang X, Liu X, Zhang H, Qiu Y (2013) Exploring techniques for vision based human activity recognition: methods, systems, and evaluation. *Sensors (Switzerland)* 13(2):1635–1650. <https://doi.org/10.3390/s130201635>
- Yan H, Zhang Y, Wang Y, Xu K (2020) WiAct: a passive WiFi-based human activity recognition system. *IEEE Sens J* 20(1):296–305. <https://doi.org/10.1109/JSEN.2019.2938245>
- Yan S, Xiong Y, Lin D (2018) Spatial temporal graph convolutional networks for skeleton-based action recognition, arXiv, 2018
- Yao L et al (2018) Compressive representation for device-free activity recognition with passive RFID signal strength. *IEEE Trans Mob Comput* 17(2):293–306. <https://doi.org/10.1109/TMC.2017.2706282>
- Yao S., Hu S, Zhao Y, Zhang A, Abdelzaher T (2017) DeepSense: A unified deep learning framework for time-series mobile sensing data processing. In: *26th international world wide web conferences WWW 2017*, pp 351–360. <https://doi.org/10.1145/3038912.3052577>
- Yao S et al. (2019) SADeepSense: self-attention deep learning framework for heterogeneous on-device sensors in internet of things applications. In: *Proceedings of IEEE INFOCOM*, vol 2019-April, pp 1243–1251. <https://doi.org/10.1109/INFOCOM.2019.8737500>
- Yao S et al. (2018) Cover feature embedded deep learning, 2018, [Online]. <https://fardapaper.ir/mohavaha/uploads/2018/06/Fardapaper-Deep-Learning-for-the-Internet-of-Things.pdf>
- Zeng M et al. (2015) Convolutional Neural Networks for human activity recognition using mobile sensors. In: *Proceedings of 2014 6th international conference on mobile computing, applications and services MobiCASE 2014*, vol 6, pp 197–205, 2015. <https://doi.org/10.4108/icst.mobicase.2014.257786>
- Zhang H, Parker LE (2016) CoDe4D: color-depth local spatio-temporal features for human activity recognition from RGB-D videos. *IEEE Trans Circuits Syst Video Technol* 26(3):541–555. <https://doi.org/10.1109/TCSVT.2014.2376139>
- Zhang D, Zhou J, Guo M, Cao J, Li T (2011) TASA: tag-free activity sensing using RFID tag arrays. *IEEE Trans Parallel Distrib Syst* 22(4):558–570. <https://doi.org/10.1109/TPDS.2010.118>
- Zhang M, Sawchuk AA (2012) USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: *UbiComp '12—proceedings of 2012 ACM conference on ubiquitous computing*, pp 1036–1043
- Zhou X, Liang W, Wang KIK, Wang H, Yang LT, Jin Q (2020) Deep-learning-enhanced human activity recognition for internet of healthcare things. *IEEE Internet Things J* 7(7):6429–6438. <https://doi.org/10.1109/JIOT.2020.2985082>
- Zhu R et al (2019) Efficient human activity recognition solving the confusing activities via deep ensemble learning. *IEEE Access* 7:75490–75499. <https://doi.org/10.1109/ACCESS.2019.2922104>
- Zhu C, Sheng W (2012) Realtime recognition of complex human daily activities using human motion and location data. *IEEE Trans Biomed Eng* 59(9):2422–2430. <https://doi.org/10.1109/TBME.2012.2190602>
- Zou H, Zhou Y, Arghandeh R, Spanos CJ (2019) Multiple kernel semi-representation learning with its application to device-free human activity recognition. *IEEE Internet Things J* 6(5):7670–7680. <https://doi.org/10.1109/JIOT.2019.2901927>
- van Kasteren TLM, Englebienne G, Kröse BJA (2011) Human activity recognition from wireless sensor network data: benchmark and software, pp 165–186. [https://doi.org/10.2991/978-94-91216-05-3\\_8](https://doi.org/10.2991/978-94-91216-05-3_8)

## Authors and Affiliations

**Neha Gupta**<sup>1,3</sup> · **Suneet K. Gupta**<sup>1</sup> · **Rajesh K. Pathak**<sup>2</sup> · **Vanita Jain**<sup>3</sup> · **Parisa Rashidi**<sup>4</sup> · **Jasjit S. Suri**<sup>5,6</sup> 

<sup>1</sup> CSE Department, Bennett University, Greater Noida, UP, India

<sup>2</sup> Rawatpura Sarkar University, Raipur, Chhattisgarh, India

<sup>3</sup> Bharati Vidyapeeth's College of Engineering, Paschim Vihar, New Delhi, India

<sup>4</sup> Intelligent Health Laboratory, Department of Biomedical Engineering, University of Florida, Gainesville, USA

<sup>5</sup> Stroke Diagnostic and Monitoring Division, AtheroPoint™, Roseville, CA 95661, USA

<sup>6</sup> Global Biomedical Technologies, Inc., Roseville, CA, USA