

Review

# Human Activity Recognition for Production and Logistics—A Systematic Literature Review

Christopher Reining<sup>1,\*</sup> , Friedrich Niemann<sup>1</sup> , Fernando Moya Rueda<sup>2</sup> , Gernot A. Fink<sup>2</sup>   
and Michael ten Hompel<sup>1</sup> 

<sup>1</sup> Chair of Materials Handling and Warehousing, TU Dortmund University, Joseph-von-Fraunhofer-Str. 2-4, 44227 Dortmund, Germany

<sup>2</sup> Pattern Recognition in Embedded Systems Groups, TU Dortmund University, Otto-Hahn-Str. 16, 44227 Dortmund, Germany

\* Correspondence: christopher.reining@tu-dortmund.de; Tel.: +49-231-755-3228

Received: 14 June 2019; Accepted: 19 July 2019; Published: 24 July 2019



**Abstract:** This contribution provides a systematic literature review of Human Activity Recognition for Production and Logistics. An initial list of 1243 publications that complies with predefined Inclusion Criteria was surveyed by three reviewers. Fifty-two publications that comply with the Content Criteria were analysed regarding the observed activities, sensor attachment, utilised datasets, sensor technology and the applied methods of HAR. This review is focused on applications that use marker-based Motion Capturing or Inertial Measurement Units. The analysed methods can be deployed in industrial application of Production and Logistics or transferred from related domains into this field. The findings provide an overview of the specifications of state-of-the-art HAR approaches, statistical pattern recognition and deep architectures and they outline a future road map for further research from a practitioner's perspective.

**Keywords:** Human Activity Recognition; production; Logistics; Motion Capturing; Inertial Measurement Unit; accelerometer; deep learning; statistical pattern recognition

## 1. Introduction

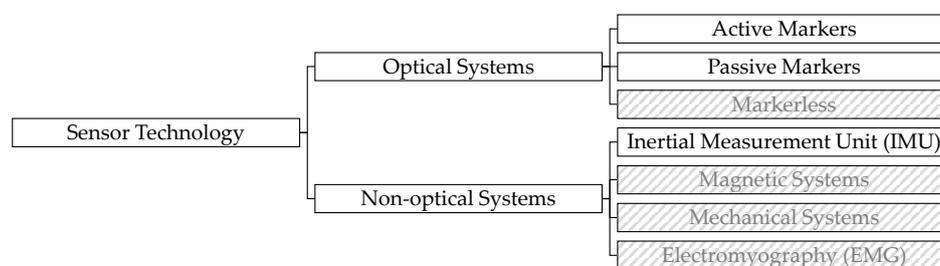
In the vision of Industry 4.0, tasks and responsibilities are shared among human employees and robots [1,2]. Nevertheless, it is expected that manual activities remain dominant in Production and Logistics (P+L) with a steady number of employees [3,4]. Automated robots are not expected to fully replace manual labour in P+L in the foreseeable future [5,6]. This is because it remains challenging to imitate the cognitive and motor skills of humans by machines [7]. Detailed information on the occurrence, duration and properties of relevant human activities is crucial to draw conclusions on how to enhance employee performance [8]. It is seen as a managerial failure not to account for human characteristics [9]. Due to advancements in sensor technology and data processing, IT-supported approaches for automated activity recognition and assessment are gaining significance [10].

Human Activity Recognition (HAR) is the task of classifying human movements. HAR methods became relevant in applications such as mobile- or ambient-assisted living, smart-homes, rehabilitation, health support, and industrial settings [11–14]. HAR commonly processes signals from videos, Motion Capturing (MoCap) systems, a set of on-body sensors or other data sources [14–17]. Traditionally, methods of statistical pattern recognition have been utilised for recognising human movements [18–20]. These methods extract relevant hand-crafted features from pre-processed and segmented sequences, and they train a classifier for assigning action labels to the sequences. Recently, deep learning methods are being used for combining the aforementioned pipeline into a single method.

They have become the state-of-the-art method for solving HAR problems in the context of gesture recognition, activities of daily living (ADL) as well as in industrial settings [11,21,22].

Automated recognition of human activities in P+L requires their definition, for example *Locomotion, Retrieval, Utility Usage* [23]. The underlying assumption is that sensor patterns can be assigned explicitly to activity classes that are known at design time and that remain identical at run time [24,25]. In industrial environments, there is no finite set of activities that can be segregated unambiguously as their definition may vary depending on the use case. As a simple example, items such as boxes or tools can be picked with the left hand, the right, or both hands and from different heights. Thinking of further activities for each variant results in an extensive quantity with overlapping features. It may be of interest to differentiate between fast and slow walking, lifting a box from the ground or from a shelf, paper-based or digital pick confirmation, and so forth. The characteristics of human motion are expected to change continuously as new information and handling technology emerges and facility layouts and processes adapt constantly.

Sensor Technology for capturing human motions is divided into optical and non-optical systems [26]. Optical systems rely on active or passive markers that are attached to a person or they are deployed markerless, using for instance RGB(-D) cameras. Optical, marker-based Motion Capturing (OMMC) is considered more precise than markerless systems, in particular for clinical studies [27,28]. On the downside, their use is restricted to constrained laboratory settings as vibration, dust, rapid temperature changes make their use in industrial environments unfeasible. Among the non-optical systems, Inertial Measurement Units (IMUs) have become highly relevant as they can be deployed in the challenging environments of P+L. They are not affected by occlusion and they do not portray human identities as in the case of videos [12]. These low-power devices are cheap, highly reliable, non-invasive and easy-to-use. In general, attaching sensors or markers to a person's body may lead to an impairment when performing manual activities. Therefore, minimalistic recording set-ups are preferred in industrial application. Magnetic and mechanical systems as well as Electromyography are further options, but they are not the focus of this contribution. The examined sensor technologies are highlighted in Figure 1 and build the foundation for the literature review.



**Figure 1.** Sensor Technology for capturing human motions [26]. The selection covered in this contribution is highlighted.

In HAR, recording, analysing, and annotating measurements from IMUs is an expensive and time-demanding task. HAR faces challenges with regards to the settings of the recording environment, number of participants as well as sensors and their configuration [20,29]. Due to the intra- and inter-class variability of human motion, a high quantity of observations from different subjects is necessary [11]. Markers and sensors may be configured with different sampling rates and resolutions. Besides, their placement on the human body may vary [30]. Therefore, data collection and annotation for HAR causes immense effort. In addition, raw inertial measurements are visually difficult to interpret by a human annotator. Additional video streams are necessary to make the observed activity apparent. This becomes a problem in real scenarios where video streams might not be allowed [23,29]. Furthermore, annotations may be inconsistent among different annotators. Additional repetitions and validations are needed to enhance the quality of the dataset while further increasing the annotation effort. Gathering data in real scenarios requires a great expense and the data are prone to be disturbed

by external factors. Nonetheless, the closeness to reality of the recorded data is ensured. On the downside, such an approach implies that the facility layout does not change until data analysis has concluded and an activity classifier has been trained. Controlled environments are appealing, as sensor measurements are less affected by noise [31,32] and recordings can be repeated under different settings and layouts that do not yet exist. Highly precise OMMC-Systems can be used in controlled environments [31]. The skeleton model visualisation of a human wearing an OMMC-Suit renders synchronised videos obsolete, facilitating the annotation process.

Considering the aforementioned points from a P+L practitioner's perspective, the following guiding questions towards the current state of research are derived:

1. What is the current status of research regarding HAR for P+L and related domains from a practitioner's perspective?
2. What are the specifications of current applications regarding the sensor technology, recording environment and utilised datasets?
3. What methods of HAR are deployed?
4. What is the research gap to enhance HAR in P+L? What does the future road map look like?

The scope of this contribution is to answer those questions by conducting a systematic literature review.

The remainder of this contribution is structured as follows. In Section 2, this contribution is demarcated from related surveys to underline its novelty value. Next, the method of the literature review is presented in Section 3. In Section 4, the findings of the review are presented and assessed. This contribution concludes with a discussion of the guiding questions in Section 5.

## 2. Demarcation from Related Surveys

In total, 68 surveys and literature reviews were identified during the review process. Among those, six were identified as having a scope related to this contribution. They are listed in Table 1 to underline the novelty value of this systematic literature review. A short content description is given in regards to the commonalities and differences to the scope of this contribution. The surveys are listed in chronological ascending order.

**Table 1.** Related Surveys.

Ref.	Year	Author & Description
[33]	2013	Lara and Labrador reviewed the state of the art in HAR based on wearable sensors. They addressed the general structure of HAR systems and design issues. Twenty-eight systems are evaluated in terms of recognition performance, energy consumption and other criteria.
[34]	2014	Xing Su et al. surveyed recent advances in HAR with smartphone sensors and address experiment settings. They divided activities into five types: living, working, health, simple and complex.
[35]	2015	Attal et al. reviewed classification techniques for HAR using accelerometers. They provided an overview of sensor placement, detected activities and performance metrics of current state-of-the-art approaches.
[36]	2016	Edwards et al. presented a review on publicly available datasets for HAR. The examined sensor technology includes MoCap and IMUs. The observed application domains are ADL, surveillance, sports and generic activities, meaning that a wide variety of actions is covered.
[20]	2018	Twomey et al. surveyed the state-of-the-art in activity recognition using accelerometers. They focused on ADL and examined, among other issues, the sensor placement and its influence on the recognition performance.
[37]	2018	O'Reilly et al. synthesised and evaluated studies which investigate the capacity for IMUs to assess movement quality in lower limb exercises. The studies are categorised into three groups: exercise detection, movement classification or measurement validation.

While all related surveys deal with HAR based on IMUs or OMMC data, P+L is not addressed. Therefore, the selection of reviewed approaches was not performed in regards to the specific demands of P+L as outlined in Section 1.

### 3. Method of Literature Review

This literature review is based on the guidelines suggested by Kitchenham and Brereton [38], Kitchenham et al. [39], Kitchenham [40] and Chen et al. [41]. The three-step pipeline of the method is illustrated in Figure 2. Based on Inclusion Criteria, a list of potentially relevant publications was created. During the selection process, three reviewers (C.R., F.N. and F.M.R.), who are experts from the domains of P+L and HAR, assigned the listed publications to one of the four stages according to predefined Content Criteria. Publications that reached Stage IV were considered relevant for the literature analysis. Their specifications were documented by the reviewers in a structured manner and they served as a point of departure for further literature search.

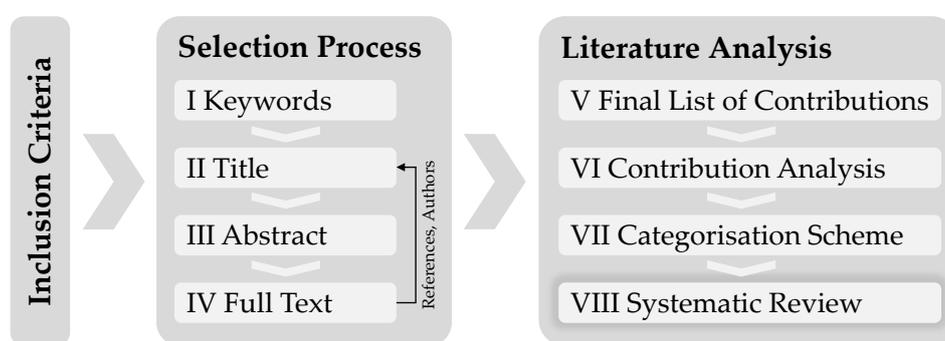


Figure 2. Method of the literature review.

The three steps of the method are explained in greater detail in the remainder of this section.

#### 3.1. Inclusion Criteria

First, criteria for including literature were defined (see Table 2). As this contribution targets the current state of the art, original contributions are of primary interest. Surveys and literature reviews were added to a separate list (see Section 2). Nevertheless, the same Inclusion Criteria applied to both types of contribution.

Table 2. Inclusion Criteria for original contributions and related surveys.

Inclusion Criteria	Description
Database	IEEE Xplore, Science Direct, Google Scholar, Scopus, European Union Digital Library (EUDL), ACM Digital Library, LearnTechLib, Springer Link, Wiley Online Library, dblp computer science bibliography, IOP Science, World Scientific, Multidisciplinary Digital Publishing Institute (MDPI), SciTePress Digital Library (Science and Technology Publications)
Keywords	Motion Capturing, Motion Capture, MoCap, OMC, OMMC Inertial Measurement Unit, IMU, Accelerometer, body-worn/on-body/wearable/wireless Sensor (Human) Activity/Action, Recognition, HAR Production, Manufacturing, Logistics, Warehousing, Order Picking
Year of publication	2009–2018
Language	English
Source Types	Conference Proceedings & Peer-reviewed Journals
Identifier	Persistent Identifier mandatory (DOI, ISBN, ISSN, arxiv)

The literature search was conducted on the selected databases and according to feasible combinations of the listed keywords. They were defined by the reviewers. Synonyms and spelling variations of the keywords were also considered. All contributions must be published between 1 January 2009 and 31 December 2018. Going further back in time would not be expedient to capture the state of the art in a rapidly progressing field of research such as HAR. Only English publications were considered. Accepted source types were conference proceedings and peer-reviewed journals. Grey literature, such as technical reports, work in progress, homepages and student degree theses, were excluded from the reviewing process to ensure the quality of the observed material.

The creation of the initial literature list to consider during the selection process happened without interference by reviewers as the Inclusion Criteria were objective. A duplicate removal took place once all databases were searched.

### 3.2. Selection Process

The selection process was conducted for all potentially relevant publications that meet the Inclusion Criteria. Each document was reviewed by each of three reviewers independently. When this subjective process led to disagreement, the Content Criteria were discussed and possibly adjusted until agreement was reached. Ultimately, the Content Criteria in Table 3 were applied for the selection process.

**Table 3.** Content Criteria for selection process.

Content Criteria	Description
(A) IMU or OMMC	Method is based on data from IMUs or OMMC-Systems. The sensors and markers are either attached to the subject's body or body-worn.
(B) Human	Contribution addresses the recognition of activities performed by humans.
(C) Physical World	Data are recorded in the physical world without the use of simulated or immersive environments.
(D) Quantification	The application aims to quantitatively determine the occurrence of activities, not to capture and analyse them for developing new methods in related fields.
(E) Application-oriented	Perspectives for deploying the proposed method in P+L is conceivable. Definition of HAR-related terms is not the contribution's focus.
(F) Physical activity	According to Caspersen et al., [42] " <i>physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure</i> ". In this literature review, bodily movement is limited to torso and limb movement.
(G) No focus on hardware	Comparison of sensor technologies or a showcase of new hardware when using it for HAR is not the contribution's focus.
(H) Clear Method	Publications are computer science oriented, stating clear pattern recognition methods and performance metrics.

Table 4 explains the four stages that contributions could reach during the selection process in accordance with the Content Criteria.

In cases where the content of a peer-reviewed journal paper and an earlier conference publication overlapped, the journal paper was preferred. The same procedure was applied when an author wrote several papers with the same scope, refining the applied methods. In these cases, solely the newest contribution was considered. Original contributions that reached Stage IV and related surveys (see Section 2) served as a starting point for further literature search. References from the Stage IV publications and other contributions from the corresponding authors were examined according to the stages of the selection process, starting with the title. The Inclusion and Content Criteria remained the same. After examination of the reference lists and the authors' profiles, a second duplicate removal took place.

**Table 4.** Stages I–IV of selection process.

Stage	Description
(I) Keywords	Keywords of the publication match with the Inclusion Criteria. Contributions have not yet been examined by the reviewers at this point.
(II) Title	The title does not conflict with any Content Criteria. This is because the title either complies with the criteria or it is ambiguous.
(III) Abstract	The abstract's content does not conflict with any Content Criteria. This is because the content either complies with the criteria or necessary specifications are missing.
(IV) Full Text	Reading the full text confirms compliance with all Content Criteria. Properties of the publication are recorded in the literature overview.

### 3.3. Literature Analysis

Once the selection process was complete, a systematic literature analysis was conducted for all contributions assigned to the fourth stage.

In Stage V, the final list of relevant contributions was extended by including the authors' names and affiliation, year of publication, name of journal or conference and the Field Weighted Citation Impact (FWCI): "[...] FWCI is an indicator of mean citation impact, and compares the actual number of citations received by a document with the expected number of citations for documents of the same document type (article, review, book, or conference proceeding), publication year, and subject area" [43]. Therefore, recording the FWCI is expected to give valuable insights on the contribution's relevance for their respective domain. The metrics were taken from the Scopus Database [www.scopus.com/sources](http://www.scopus.com/sources) on 6 June 2019.

Once the general information was acquired, each contribution's striking features in regards to this review's guiding questions were analysed. Thus, Stage VI required the reviewers to read all relevant literature and briefly summarise:

- the initial situation and scope;
- the methodological and empirical results; and
- the further research demand.

This procedure enabled the reviewers to get an overview of the research area. Beyond that, the application domain, the sensor technology and utilised methods and datasets were given attention when studying the literature.

Based on the contribution analysis, a categorisation scheme was derived by the reviewers in Stage VII. This scheme consists of two layer–root categories and subcategories. The initial categories were created by the reviewers. As in the selection process, points of disagreement were discussed among the reviewers. The initial categories were subject to change when allocating the contributions in the following stage.

The systematic review of the literature was accompanied by a continuous refinement of the categorisation scheme—ultimately leading to the one illustrated in Table 5. After the categorisation scheme was finalised, all publications were allocated accordingly in Stage VIII. Per root category, the criteria for multiple subcategories can be met, e.g., a contribution can consider both working activities and exercises or attach sensors and markers to different body parts at the same time.

Apart from categories in Table 5, the evaluation metrics of HAR were also analysed. As HAR typically contains imbalanced datasets, proper metrics for comparing the performance of different methods have to be chosen. In general, the accuracy is used extensively for evaluating algorithms for solving classification tasks. Nevertheless, the F1-measures are more convenient for evaluating the performance of HAR using highly imbalanced datasets. The mean F1 measure is the harmonic mean of the precision and recall. Additionally, the weighted F1 measure considers the proportion of the class in the dataset.

**Table 5.** Categorisation scheme.

<b>Root Category</b>	
Subcategory	Description
<b>Domain</b>	
P+L	Deployment in industrial settings, e.g., production facilities or warehouses
Other	Related application domain, e.g., health or ADL
<b>Activity</b>	
Work	Working activities such as assembly or order picking
Exercises	Sport Activities, e.g., riding a stationary bicycle or gymnastic exercises
Locomotion	Walking, running as well as the recognition of the lack of locomotion when standing
ADL	Activities of daily living including cooking, doing the laundry, driving a car and so forth
<b>Attachment</b>	
Arm	Upper and lower arm
Hand	including wrists
Leg	including knee and shank
Foot	including ankle
Torso	including chest, back, belt and waist
Head	including sensors attached to a helmet or protective gear
Smartphone	Worn in a pocket or a bag. If attached to a limb, the subcategory is checked as well
<b>Dataset</b>	
Repository	Utilised dataset is available in a repository
Individual	Dataset is created specifically for the contribution and not available in a repository
Laboratory	Recording takes place in a constraint laboratory environment
Real-life	Recording takes place in a real-life environment, e.g., a real warehouse or in public places
Name of dataset	Name, origin, repository and description of dataset
<b>Sensor</b>	
Passive Markers	Markers reflect light for the camera to capture
Active Markers	Markers emit light for the camera to capture
IMU	Devices that measure specific forces such as acceleration or gyroscopes
<b>Data Preparation (DP)</b>	
Pre-Pr.	Pre-Processing: Normalisation, noise filtering, low-pass and high-pass filtering, and re-sampling
Segm.	Segmentation: Sliding window-approach
<b>Shallow Method</b>	
FE - Stat. Feat.	Statistical feature extraction: Time- and Frequency-Domain Features
FE- App.-based	Application-based features, e.g., Kinematics, Body model, Event-Based
FR	Feature reduction, e.g., Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA), Kernel Discriminant Analysis (KDA), Random Projection (RP)
CL-NB	Classification method: Naïve Bayes
CL-HMMs	Classification method: Hidden Markov Models
CL-SVM	Classification method: Support Vector Machines
CL-MLP	Classification method: Multilayer Perceptron
CL-Other	Classification method: Random Forest (RF), Decision Trees (DT), Dynamic Time Warping (DTW), K-Nearest Neighbor (KNN), Fuzzy-Logic (FL), Logistic Regression (LR), Bayesian Network (BN), Least-Squares (LS), Conditional Random Field (CRF), Factorial Conditional Random Field (FCR), Conditional Clauses (CC), Gaussian Mixture Models (GMM), Template Matching (TM), Dynamic Bayesian Mixture Model (DBMM), Emerging Patterns (EP), Gradient-Boosted Trees (GBT), Sparsity Concentration Index (SCI)
<b>Deep Learning (DL)</b>	
CNN	Convolutional Neural Networks
tCNN	Temporal CNNs and Dilated tCNNs (DTCNN)
rCNN	Recurrent Neural Networks, e.g., GRU, LSTM, Bidirectional LSTM

#### 4. Results

Following the method as outlined in Section 3, the results of the literature review are presented in this section. First, the results of the selection process are illustrated and major causes for paper exclusion in accordance with the Content Criteria for selection process are pointed out. This is followed by a structural analysis of the Stage IV literature using the predefined categorisation scheme.

#### 4.1. Contributions Per Stage and Reasons for Exclusion during Selection Process

During selection process, 1243 publications were examined as they complied with the Inclusion Criteria, with 52 of them making it to Stage IV as their full text complied with the Content Criteria for selection process. Table 6 illustrates the number of contributions per stage.

**Table 6.** Examined Publications per Stage.

Stage	No. of Publications
(I) Keywords	1243
(II) Title	524
(III) Abstract	263
(IV) Full Text	52

Referring to the predefined Content Criteria for selection process, common cases are presented where contributions violated the criteria during selection process.

Regarding Criterion A, two major issues kept contributions from reaching Stage IV. Approaches of vision-based activity recognition using images and videos have been very prominent in the literature corpus, even though they were not explicitly searched for based on the inclusion criteria keyword list. Beyond that, contributions were excluded despite using IMUs, because the sensors were not attached on the human body but to objects such as golf clubs, surfboards, ice hockey sticks, skateboards and so forth.

The keywords *IMU* and *accelerometer* led to contributions dealing with sensor-based control of robots and unmanned aerial vehicles (UAVs). Another active field of research is activity recognition of animals, e.g., dogs, horses and lizards. Both issues meant a violation of Content Criteria B.

In contradiction with regards to Content Criterion C, many contributions do not gather data in the physical world. Instead, motion patterns are taken from a 3D-simulation of a human working in a digital twin of a scenario resembling a P+L facility. Researchers also use virtual reality to create a feeling of immersion and record data of a human working in this simulated environment. Thus far, there is no empirical proof known to the authors that motion patterns recorded from a human linked to a virtual reality can be used for HAR in a real-world facility.

There are fields of research related to HAR that do not deal with quantification of human activity (Content Criterion D). Rather, they aim at qualitative assessment, e.g., in regards to ergonomics. Further examples are robot control via gestures and the analysis of human behaviour to create more realistic digital human simulation models. An active field of research related to HAR is emerging in the domain of medicine. Advances in sensor technology sparked research aiming at the detection of clinical pictures and their consequences, e.g., Parkinson's disease and Alzheimer's disease, stroke, medial knee osteoarthritis, leg prostheses, asymmetries during the gait cycle, head trauma and many more.

Contributions that did not provide a use case and thus cannot be considered application-oriented, violated Content Criterion E. They were either a survey and thus moved to a separate list (see Section 2), or they were dealing with fundamentals of HAR, lacking the prospect of deployment in P+L from a practitioner's perspective. In the case the application is not taking place in P+L, the reviewers discussed whether its transfer to this domain seems feasible. While locomotion and handling activities that are part of daily living, e.g., using tools, seemed transferable, there were cases where this was not the case. The first group of these activities were sport activities such as jumping and dribbling in basketball, boxing, alpine skiing, soccer, ball throwing in baseball and handball, and underwater sports, e.g., fly kicks. The second group consists of ADLs, that do not resemble P+L activities. Most present among them were dance moves, performing music with a violin or piano, or locomotion with a rolling walker, crutches or a wheelchair.

In the literature, the meaning of the term *physical activity* depends on the context. Consequently, the search for *Activity* or *Human Activity* often resulted in violations of Content Criterion F. Eye tracking and the recognition of facial expressions and emotions such as stress cannot be considered limb or

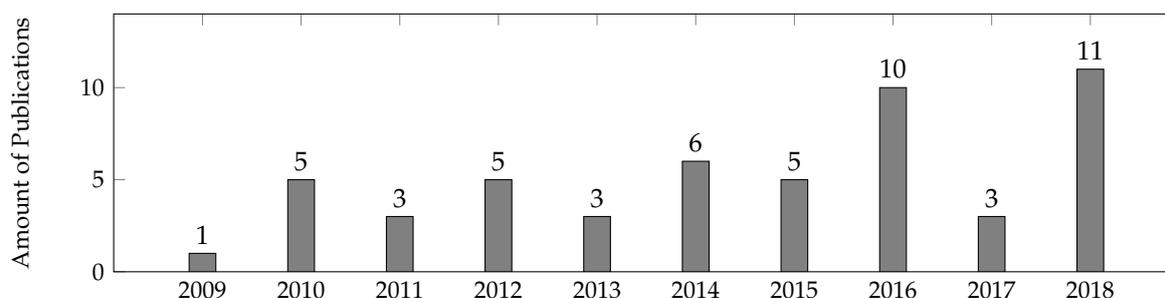
torso movements. Work on pose tracking and path analysis did not aim to recognise the physical activity. For example, it is not necessarily the case that the employee moved on foot as he/she could have used a vehicle.

Advances in sensor technology for HAR applications are published in journals and conference proceedings. Current trends are devices for finger, head and back movement tracking that can be deployed easily. These hardware showcases may be useful for HAR in the long run, but violate Content Criterion G. Contributions that do not show practical application scenarios as they solely focus on the hardware showcase were excluded.

Contributions with no clear pattern recognition methods and non-standard performance metrics for HAR were not considered, e.g., contributions with vague procedure on the data collection, pre-processing, segmentation and classification. There are contributions that rely on conditions flow-charts based on thresholding, or they simplify the algorithms by only mentioning them.

#### 4.2. Systematic Review of Relevant Contributions

Analysing the number of relevant publications per year reveals an increase in the second half of the observed ten-year time interval. Thus, the growing relevance of this review's scope is confirmed (see Figure 3).



**Figure 3.** Papers distribution by year of publication.

The peaks in the years 2016 and 2018, in particular when compared to 2017, cannot be explained by special issues of domain-specific journals. The 52 publications are spread over 41 different journals or conferences. This fact underlines that HAR for P+L benefits from a wide range of interdisciplinary expertise. The International Workshop on Sensor-based Activity Recognition and Interaction (iWOAR) accounts for three publications, making it the most frequent one. The remaining journals or conferences are accountable for a single or two contributions each. In total, 165 unique authors were involved in writing the 52 contributions. Their countries of affiliation are spread worldwide over 26 countries. None of the authors has participated in writing more than three publications.

The results of the systematic literature review according to the predefined Categorisation Scheme are illustrated in Table 7. Entries are listed in chronological ascending order and alphabetically according to the last name of the corresponding author.



Table 7. Cont.

General Information			Domain		Activity		Attachment					Dataset				DP		Shallow Method							DL										
Ref.	Year	Author	FWCI	P+L	Other	Work	Exercises	Locomotion	ADL	Arm	Hand	Leg	Foot	Torso	Head	Smartphone	Repository	Individual	Laboratory	Real-Life	Pre-Pr.	Segm.	FE-Stat.Feat.	FE-Others	FR	CL-NB	CL-HMMs	CL-SVM	CL-MLP	CL-Others	CNN	tCNN	rCNN		
[79]	2017	Song-Mi Lee et al.	10.91		x		x								x		x		x		x												x		
[80]	2017	Scheurer et al.	7.06		x		x							x			x	x		x		x						x					GBT, KNN		
[81]	2017	Vital et al.	0.93		x		x	x	x	x	x	x	x	x				x	x					x	x	x		x	x				DBMM		
[82]	2018	Chen et al.	1.77		x		x							x		x	x		x			x						x					x		
[83]	2018	Moya Rueda et al.	3.69	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x	x	x												x	
[84]	2018	Nair et al.	0.00		x		x							x		x	x		x	x	x	x												x	
[85]	2018	Reining et al.	0.00	x		x				x	x	x	x	x	x				x	x		x												x	
[86]	2018	Tao et al.	0.00	x		x				x									x	x		x	x											x	
[87]	2018	Wolf et al.	0.00		x		x	x							x				x	x		x												x	
[88]	2018	Xi et al.	1.61		x		x	x	x	x	x	x	x	x			x		x	x		x												x	
[89]	2018	Xie et al.	6.43		x		x							x				x	x			x	x					x						RF	
[22]	2018	Yao et al.	3.53		x		x	x	x	x	x	x	x	x			x	x	x	x	x														x
[90]	2018	Zhao and Obonyo	0.00	x		x		x		x		x		x	x				x	x		x	x		x	x									KNN
[91]	2018	Zhu et al.	0.00		x		x							x		x	x		x			x	x				x								RF,KNN,LR
<b>Total: 52</b>				<b>8</b>	<b>47</b>	<b>8</b>	<b>13</b>	<b>46</b>	<b>23</b>	<b>19</b>	<b>24</b>	<b>16</b>	<b>15</b>	<b>35</b>	<b>6</b>	<b>21</b>	<b>18</b>	<b>38</b>	<b>19</b>	<b>39</b>	<b>20</b>	<b>40</b>	<b>32</b>	<b>7</b>	<b>8</b>	<b>11</b>	<b>7</b>	<b>16</b>	<b>10</b>				<b>7</b>	<b>7</b>	<b>4</b>

During categorisation, it turned out that active markers are not used in any of the relevant contributions. OMMC is used in only two contributions [54,85]. The remaining research was done using IMUs, sometimes in conjunction with other sensor technology that is not considered in this review. Therefore, the utilised sensor technology is not explicitly stated in Table 7.

The FWCI could not be recorded for three publications [48,57,71]. They are marked with a “-”. Some of the publications from 2018 were published at the end of the year. Thus, it is too early to draw conclusions from their FWCI of 0. The highest FWCI of 64.63 is held by Bulling et al. The mean value of the remaining 49 publications—including those with a FWCI of 0 but excluding those where it could not be captured—is 10.04.

In the following, the major findings per category are summarised in two parts—Application and HAR Methods.

#### 4.2.1. Application

The first part of the systematic review focuses on the application domain, the observed activities, the sensor attachment and the utilised datasets. When pointing out the major findings, particular emphasis is put on the contributions from the P+L domain.

##### Domain

Eight publications are assigned to the P+L domain, three of them in conjunction with other domains. Interestingly, all publications of the P+L domain have been published since 2013. This underlines the increasing significance and research interest in HAR for P+L. The application covers a wide range of sectors, such as manufacturing and assembling [58,86], warehousing [72,83,85], construction [90] and maintenance [11,14]. Warehousing, in particular order picking, is the only sector in logistics. Other areas of logistics, such as packaging, in-house transport, external transport and handling processes such as loading, unloading and reloading are not observed in the contributions. The majority of the remaining 44 contributions focus on simplistic activities or simple exercises, as pointed out in the next paragraph.

##### Activity

The eight contributions from the P+L domain correspond to those that cover working activities. Tao et al. and Koskimaki et al. focused solely on working activities but did not involve locomotion. Zhang and sawchuk did not recognise the activity itself but motions that are performed during working, such as neck bending, neck extension, kneeling and so forth. The Skoda dataset utilised in [11,14] involves manipulative gestures in car maintenance, but does provide labelled data of locomotion (see Section 4.2.1.4). However, the other datasets utilised in these contributions do involve locomotion activities. The remaining publications [72,76,83] from P+L apply HAR methods on both stationary working processes and locomotion activities. Apart from Reining et al. [85], the observed contributions in the domain of P+L assume that the activity definition is known at design time and is not going to change at run time of a HAR method (see Section 1). Attribute-based representations are proposed to address this issue, following the concept of Rueda and Fink [92] and Lampert et al. [93].

Within the entire literature corpus, locomotion is most often represented by far. Forty-six contributions aim to recognise the associated activities such as walking, running, standing and climbing stairs that can be recorded without elaborate set-ups. In total, 23 contributions deal with ADL, followed by exercises with 13 occurrences. It is noticeable that more complex activities that require greater effort for recording are represented less than simple ones. This is because the method of HAR is the focus of most publications, while datasets are often regarded as a tool for evaluation. The domain in which the data were created is of secondary importance.

## Attachment

Among the contributions on P+L, passive MoCap sensor technology is applied only once [85]. The remaining contributions from P+L utilise IMUs. The combination of Surface Electromyography (sEMG) and IMUs was proposed by Tao et al. Using a single IMU attached to a hand or arm is proposed twice [58,86].

Examining the sensor attachment in the entire literature corpus, no clear picture emerges. While placing the sensors on the torso is most prominent (35), other locations are common as well. The reviewers could not establish a link between the sensor placement and the recognised activities. While some researchers attach the sensors to the arm performing a motion, e.g., drilling or opening a window, other approaches try to recognise ADLs and locomotion activities with sensors placed on the waist. The same issue also occurs the other way around, when researchers aim to recognise locomotion with IMUs placed on the hand. Out of the 21 contributions utilising smartphone sensors, 13 include the creation of individual datasets. The latter group solely takes smartphone data into consideration without deploying further sensors. Placing sensors on the head or a helmet is proposed rather rarely. Among the six contributions that do so, Wolff et al. presented the only work based on the idea of using head-worn sensors exclusively without further body-worn devices. Transferring this approach to P+L could possibly minimise the impairment of employees when performing manual work.

## Dataset

Seven out of the eight P+L contributions use data recorded in a laboratory. Work stations or relevant features of a facility are rebuilt to record data close to reality. Nevertheless, these data are not recorded in a real process with those employees that routinely perform the manual activities. There are four P+L contributions that utilise data recorded from real-life working routines, of which three also use laboratory data. There is a single P+L contribution that solely uses real-life data [72]. No work could be found on the issue of training a classifier using data from a laboratory for deployment in a real-life P+L facility. Besides, no work was found concerning transfer learning between datasets and scenarios in an industrial context. Examining the entire literature corpus, there are only six papers that use both real-life and laboratory data versus 39 real-life and 19 laboratory data, respectively.

There are six P+L contributions using individual datasets, while data from a repository are utilised three times. A single P+L contribution uses both data sources [83]. Examining the entire literature corpus, it is striking that a majority of 38 contributions utilises individual datasets, and 34 of them do so without referring to data that are publicly available in repositories. Thus, there are four contributions using both individual and repository data [22,54,71,83]. As most contributions lack a recording protocol and a detailed description of the recorded activities, it remains unclear whether the motion patterns and their assigned activity labels as well as the underlying activity definitions are comparable.

During the review process, the authors traced information about the used datasets back to their origin, e.g., in a repository or the first time they were mentioned in a contribution. Table 8 lists publicly available datasets utilised by a single contribution. They are listed in alphabetical order, providing a reference for further reading and stating the contribution that utilises this dataset.

**Table 8.** Overview of publicly available datasets utilised by a single relevant contribution.

Ref.	Name	Utl. in
[94]	Actitracker from Wireless Sensor Data Mining (WISDM)	[14]
[95]	Activity Prediction from Wireless Sensor Data Mining (WISDM)	[69]
[96]	Daphnet Gait dataset (DG)	[73]
[97]	Mocap Database HDM05	[54]
[98]	Realistic Sensor Displacement Benchmark Dataset (REALDISP)	[76]
[99]	Smartphone-Based Recognition of Human Activities and Postural Transitions Dataset	[76]
[100]	USC-SIPI Human Activity Dataset (USC-HAD)	[78]
[101]	Wearable Action Recognition Database (WARD)	[68]

Table 9 describes those datasets that are used by more than one contribution in greater detail.

**Table 9.** Overview of publicly available datasets utilised by two or more relevant contributions.

Ref.	Name	Description	Utl. in
[102]	Opportunity	Published in 2012, this dataset contains recordings from wearable, object, and ambient sensors in a room simulating a studio flat. Four subjects were asked to perform early morning cleanup and breakfast activities.	[11,14,22,73,74,83,88]
[103]	Human Activity Recognition Using Smartphones Data Set	The dataset from 2012 contains smartphone-recordings. 30 subjects at the age of 19 to 48 performed six different locomotion activities wearing a smartphone on the waist.	[71,78,82,84,91]
[104]	2 PAMAP	Published in 2012, this dataset provides recordings from three IMUs and a heart rate monitor. Nine subjects performed twelve different household, sports and daily living activities. Some subjects performed further optional activities.	[66,73,76,83,88]
[105]	Hand Gesture	The dataset from 2013 contains 70 minutes of arm movements per subject from eight ADLs as well as from playing tennis. Two recorded subjects were equipped with three IMUs on the right hand and arm.	[22,29,71]
[106]	Skoda	This dataset from the year 2008 contains ten manipulative gestures performed by a single worker in a car maintenance scenario. 20 accelerometers were used for recording.	[11,14]

It was noticed that the oldest dataset from 2008, Skoda, is still used eight years later in 2016 in [11]. Even though several contributions use excerpts from the same datasets, a performance comparison between the proposed methods is not necessarily possible. In some cases, the excerpts taken from a dataset differ. One reason is that authors may exclude activities with too few subjects or solely rely on IMU data without taking other data sources into account. Since a wide variety of data is used, a comparison of classification performance seems hardly possible from a practitioner's perspective.

Irrespective of the HAR method applied, effort for dataset creation is another factor to consider in commercial applications for P+L. However, none of the authors state the effort for dataset creation. It remains unclear how much time the set-up of equipment and the recording sessions consume. The same applies to the process of annotation, in some contributions referred to as labelling. Cross tests and repetition tests may reveal the consistency of labels among several annotators and the inherent annotation error caused by vague class descriptions. In the observed contributions, the time spent on annotating and the consistency of the labels are not discussed.

#### 4.2.2. HAR Methods

The HAR methods represent the second part of the systematic review, covering the standard pattern recognition chain—this involves the pre-processing, segmentation, and shallow classification methods—and deep learning, in addition to the performance metrics.

#### Data Representation

Different data representations have been used by the publications in Table 7, depending on the HAR applications and sensors. Inertial measurements recorded by sensors integrated in IMUs are deployed, in general, for solving HAR. Usually, more than three devices are set on the human body, e.g., on the hands, legs, head, and torso. Differently, the authors of [44,82,91] recorded acceleration measurements from only one device, which is placed on the waist. The authors of [52,75,79] proposed using the magnitude of the acceleration vector from the three components  $x$ ,  $y$ , and  $z$ . The authors of [86] used the logarithm magnitude of a two-dimensional Discrete Fourier Transform of IMU signals. They proposed utilising this magnitude as an image input for a CNN.

Otherwise, the local human joint poses have also been used for HAR. In this case, Optical Motion Capturing (OMOCAP) or devices such as Kinect were deployed. In [54], the authors represented



### Shallow Methods

In addition to segmentation, a traditional supervised-HAR pipeline includes the extraction of relevant handcrafted-features, a feature reduction, a training stage and a classification. The handcrafted features should capture the intrinsic characteristics of a certain human action. A feature reduction is deployed for reducing the dimensionality of the feature space, keeping the discriminant properties of the features. In the training stage, a classifier is trained by using the extracted features and the ground-truth activity labels. Finally, the classifier assigns activity classes.

**Feature Extraction.** In the standard pattern recognition methods, the feature extraction is an important stage. It allows representing data in a compact manner, which helps with later classification stages. They are divided into two main groups, statistical features and application-based.

Table 11 shows the statistical features that are mentioned in the publications in Table 7. Time-domain features focuses on the waveform characteristics, and frequency-domain features focus on the periodic structure of the signal [90]. The Fourier transform is applied to the raw- or pre-processed signals to acquire the estimated spectral density of the time series. As a remark, the authors of [64] mentioned the usage of time- and frequency-domain features, but they did not specify which ones exactly.

**Table 11.** Statistical features divided in two main groups: time and frequency domain.

Domain	Features	Definitions	Publications
Time	Variance	Arithmetic Variance	[29,44–46,48–50,53,55–57,57,59–63,65,67–69,71,72,75,76,78,80–82,89–91]
	Mean	Arithmetic Mean	[29,45,48–50,53,55,56,59,61–63,65,67–69,71,72,75,76,78,80–82,89–91]
	Pairwise Correlation	Correlation between every pair of axes	[50,53,55,56,60–62,65,69,78,80,82,89,91]
	Minimum	Smallest value in the window	[45,46,57,69,72,76,82,90,91]
	Maximun	Largest value in the window	[45,46,57,69,72,76,82,90,91]
	Energy	Average sum of squares	[49,50,71,75,76,78,82,91]
	Signal Magnitude Area		[47,50,53,78,80,82,91]
	IQR	Interquartile Range	[55,76,78,82,89,91]
	Root Mean Square	Square root of the arithmetic mean	[50,55,56,61,75,90]
	Kurtosis		[45,46,60,68,80]
	Skewness		[45,46,50,75,80]
	MinMax	Difference between the Maximum and the Minimum in the window	[50,61,63,75]

Table 11. Cont.

Domain	Features	Definitions	Publications
Time	Zero Crossing Rate	Rate of the changes of the sign	[29,50,69]
	Average Absolute Deviation	Mean absolute deviations from a central point	[55,90]
	MAD	Median Absolute Deviation	[55,76]
	Mean Crossing Rate		[29,50]
	Slope	Sen's slope for a series of data	[90]
	Log-Covariance		[81]
	Norm	Euclidean Norm	[72]
	APF	Average Number of occurrences of Peaks	[61]
	Variance Peak Frequency	Variance of APF	[61]
	Correlation Person Coefficient		[76]
	Angle	Angle between mean signal and vector	[76]
	Time Between Peaks	Time [ms] between peaks	[48]
	Binned Distribution	Quantisation of the difference between the Maximum and the Minimum	[48]
	Median	Middle value in the window	[50]
	Five different Percentiles	Observations in five different percentiles [10, 25, 50, 75, 90]	[57]
Sum and Square Sum in Percentiles	Sum and Square sum of observations above/below certain percentile [5, 10, 25, 75, 90, 95]	[57]	
ADM	Average Derivate of the Magnitude	[62]	
Frequency	Entropy	Normalised information entropy of the discrete FFT component magnitudes of the signal	[29,44,49,53,60,63,68,71,75,80,90]
	Signal Energy	Sum squared signal amplitude	[29,56,63,68,71,82,89–91]
	Skewness	Symmetric of distribution	[76,82,90,91]
	Kurtosis	Heavy tail of the distribution	[76,82,90,91]
	DC Component of FFT and DCT		[50,67,89]
	Peaks of the DFT	First 5 Peaks of the FFT	[45,46]
	Spectral		[90]

Table 11. Cont.

Domain	Features	Definitions	Publications
Frequency	Spectral centroid	Centroid of a given spectrum	[90]
	Frequency Range Power	Sum of absolute amplitude of the signal	[90]
	Cepstral coefficients	Mel-Frequency Cepstral Coefficients	[29]
	Correlation		[49]
	maxFreqInd	Largest Frequency Component	[76]
	MeanFreq	Frequency Signal Weighted Average	[76]
	Energy Band	Spectral Energy of a Frequency Band $[a, b]$	[76]
	PPF	Peak Power Frequency	[80]

Application-based features refer to features that were created for a certain application or dataset. These features are based on geometric, structure and kinematic relations. Table 12 shows the application-based features used in Table 7.

Table 12. Application-based features for HAR.

Domain	Features	Definitions	Publications
Spatial	Gravity variation	Gravity acceleration computed using the harmonic mean of the acceleration along the three axes (x,y,z)	[65]
	Eigenvalues of Dominant Directions		[50]
Structural	Trend		[55,56]
	Magnitude of change		[55,56]
Time	Autoregressive Coefficients		[47]
Kinematics	User steps frequency	Number of detected steps per unit time	[65]
	Walking Elevation	Correlation between the acceleration along the y-axis vs. the gravity acceleration or acceleration along the z-axis	[65]
	Correlation Hand and foot	Acceleration correlation between wrist and ankle	[65]
	Heel Strike Force	Mean and variance of the Heel Strike Force, which is computed using dynamics	[65]
	Average Velocity	Integral of the acceleration	[50]

**Feature Reduction.** The authors of [44–46,68,81] deployed Principal Component Analysis (PCA) for reducing the dimensionality of their features. PCA is a holistic method that considers its inputs as points in a high-dimensional space and it finds a lower-dimensional feature space along the highest variance, where classification becomes easier. Linear Discriminant Analysis (LDA) is

another holistic method that tries to overcome the drawbacks of PCA. It minimises the intra-class variance, and maximises the inter-class variance of a set of inputs/features. It finds an optimal projection by maximising the ratio between the inter- and intra-class variations of the inputs. Kernel Discriminant Analysis (KDA) is a non-linear discriminating approach based on kernel techniques to find non-linear discriminating features, used in [29,47,47]. Quadratic Discriminant Analysis (QDA) was deployed by Siirtola and Rönning [57]. The authors of [90] followed the Recursive Feature Elimination (RFE) for finding the best set of features. The RFE can be seen as a dense parameter search, which iteratively selects or rejects a set of features after training and deploying a classifier. A Bag of Feature Representation using Kmeans for clustering was proposed by Deng et al. [54]. They used the Kmeans algorithm for clustering  $M$  motion sequences. These motion sequences are 3D joint positions and orientations through  $m$  frames. Additionally, the authors of [50,68] utilised a Random Projection (RP). Sequential forward- or backward-feature selection (SFFS) was used in [45,90].

**Classification.** The authors of [52] trained a HMM model per axis ( $x$ ,  $y$ , and  $z$ ) of pre-processed acceleration measurements, fusing them with a weighted sum. The authors of [29,62,71] also deployed HMMs, and the authors of [49] used couple HMMs. The authors of [78] used a hierarchical conditional HMMs. The authors of [44] used a NB using the Probability Density Function (PDF) from all of the 19 features, which were previously reduced using PCA; assuming that all the features in the lower space are not mutually correlated. The authors of [81,90] used also a NB assuming that the features are mutually independent. Similarly, refs. [29,45,50,56,63,68,72] used NB. The authors in [45,50,61,67,68,72,80,90] deployed SVMs. The authors of [65] used a SVM-based binary decision tree classifier. The authors of [53] proposed a hardware-friendly SVM that is meant to be deployed on smartphone devices. [29,45,46,57,63,64,68,80,90] trained a K-Nearest Neighbor (KNN) based classifier. The authors of [81] used Dynamic Bayesian Mixture Model (DBMM) for combining conditional probability outputs from different classifiers, namely, the NB, SVM and MLP. A weight is assigned to each base classifier, according to a learning process, using an uncertainty measure as a confidence level. Multilayer Perceptron (MLP)—a network with two fully connected layers and a softmax layer—was deployed by [45–48,50,56,59,61].

Additionally, other classifiers, methods or approaches were used: Dynamic Time Warping (DTW) [46,75], Conditional Random Field (CRF) [62], Random Forest (RF): [50,61,72], Decision Tree (DT) [48,56,59,68,90], Logistic Regression (LR) [48,59,61], Least Squares Method [45], Gaussian Mixture Model (GMM) [64], Template matching (TP) [75], Correlation [75], Euclidean distance (ED) [75]. In addition, the authors of [61] combined single classifiers by majority voting and average of probabilities. Joint Boosting (JB) [69], Bagging [56,69] and Stacking [69] were also deployed.

## Deep Learning

Deep architectures have recently been proposed for solving HAR. Differently from the standard HAR pipeline, deep architectures combine the feature extraction and classification in a single approach. Their features are directly learned from data, being more discriminative. Besides, they overcome some problems regarding computation and adaptability of handcrafted features.

**Convolutional Neural Networks (CNNs).** CNNs, first proposed in [107], have recently been used for solving HAR problems. CNNs combine the feature extraction and classification in an end-to-end approach. CNNs contain hierarchical structures combining convolutional operations using learnable filters and non-linear activation functions, downsampling operations, and classifiers [108]. Networks for HAR are relatively small—in comparison to networks for image classification or document analysis—with a maximum of four convolutional layers and a maximum of three pooling layers. The authors of [67] used three convolutional layers and three pooling layers. As a remark, the authors of [87] did not explain their network in detail. Classification is usually performed using a softmax layer [83,86] or by means of attribute representation, a sigmoid function and a distance measurement, e.g., Euclidean, Cosine or Bray–Curtis distances [85]. The authors of [86] used spectrograms of inertial

signals as image inputs for CNNs. Tao et al. [86] concatenated sEMG vectors, corresponding to muscles activation levels, to the output of the first fully-connected layer and input it to a softmax layer.

**Temporal Convolutional Neural Networks (tCNNs).** In multichannel time-series HAR, the CNNs input consists of a stack of segmented sequences from different sensors for a certain temporal duration. tCNNs carry out convolutions and downsampling operations along the time axis, sharing different feature extractors among sensors. The authors of [14] also used a small CNN with only one conv layer, a pooling layer and a MLP. However, they proposed parallel branches per spatial axis, x, y and z, with late fusion. The authors of [77] also used a small network with one temporal convolutional layer, one pooling and a MLP. The authors of [79] used a small tCNN with one conv layer, one pooling, one FC and a softmax layer. The authors of [83] used temporal convolutional networks. They also proposed an architecture with different parallel branches per IMU device with late fusion. The authors of [84] proposed using an Encoder–Decoder TCNN and a dilated TCNN.

**Recurrent Neural Networks (RNNs).** The authors of [11] proposed a network with four temporal convolutional layers and two Long-Short Term Memory (LSTM) layers followed by a softmax layer. The authors of [88] proposed to use dilated temporal ConvLSTMs. The network consists of one initial convolutional layer followed by three dilated temporal convolutional layers with different dilated factors. The feature maps of the last dilated temporal convolutional layer is fed to a two LSTMs layers followed by a softmax. Chen et al. [82] proposed a combination of LSTMs with raw-data input, an MLP with handcrafted-feature inputs and late fusion. Differently, Zhu et al. [91] used a LSTM network with statistical features as input.

For training the CNNs, tCNNs, and RNNs, the Adam [79,84,86] and the RMSProp [11,83–85] optimisations methods were usually used. For avoiding overfitting, the authors of [86] utilised the L2-regularisation, the authors of [23,83] added random noise to the normalized measurements, and the authors of [23,82,83,91] used dropout.

## Metrics

Following the metrics procedure from the publications in Table 7, the most used metric is the accuracy. However, following the different datasets presented in the publications, most of those datasets have an imbalance problem, that is, some classes contain more samples than other classes. Generally, the reported performances in the publications show relative good results, approaching 100% accuracy. Nevertheless, these performances must be revised. Using the F1 measures, computing the mean and weighted average of the precision and recall, could give a more impartial conclusion of the performance for HAR (see Table 13).

**Table 13.** Number of publications using these metrics as performance metric for HAR.

Metric	# of Publications
Accuracy	38
Precision	12
Recall	11
weightedF_1	5
meanF_1	6

## 5. Discussion and Conclusions

The purpose of this systematic literature review is to capture the state of the art of Human Activity Recognition for Production and Logistics. To achieve this goal, HAR applications in related domains are taken into account as well. Fifty-two contributions were selected from an initial list of 1243 publications according to predefined Inclusion and Content Criteria. The relevant literature is categorised and analysed by domain experts from P+L and HAR. Summarising the findings, the guiding questions from Section 1 are answered:

1. *What is the current status of research regarding HAR for P+L and related domains from a practitioner's perspective?*

For the past 10 years, eight publications dealing with HAR in P+L have been identified. They address a variety of use cases but none covers the entire domain. Apart from two applications [83,85], the approaches assume a predefined set of activities, which is a downside amid the versatility of human work in P+L. Furthermore, the necessary effort for dataset creation is unknown, making the expenditure for deploying HAR in industry difficult to predict. In applications for related domains, locomotion activities as well as exercises and ADLs that resemble manual work in P+L are covered, allowing for their transfer to this domain.

2. *What are the specifications of current applications regarding the sensor technology, recording environment and utilised datasets?*

The vast majority of research is done using IMUs placed on a person or using the accelerometers of smartphones. The sensor attachment could not be derived from the activities to be recognised. There was no link apparent to the reviewers. Seven out of the eight P+L contributions use data recorded in a laboratory. In total, 39 contributions use real-life data versus 19 that use laboratory data. Only four papers use both real-life and laboratory data. The reviewers did not find work regarding the training of a classifier using data from a laboratory for deployment in a real-life P+L facility or transfer learning between datasets and scenarios in an industrial context. Most of the publications proposed their own datasets or used individual excerpts from data available in repositories; thus, replicating their methods and results is hardly possible.

3. *What methods of HAR are deployed?*

Current publications solve HAR either using a standard pattern recognition algorithm or using deep networks. Publications follow only the sliding window approach for segmenting signals. The window size differs strongly according to the recording scenarios. However, the overlapping is usually 50%. For the standard methods, there is a large number of statistical features in time and frequency, being the variance, mean, correlation, energy and entropy the most common. Deep applications have been applied successfully for solving HAR. In comparison with applications in the vision domain, the networks are relatively shallow. Temporal CNNs or combinations between tCNNs and RNNs show the best results. Accuracy is the most used metric for evaluating the HAR methods. However, methods using datasets with unbalanced annotation should be evaluated with precision, recall and F1-metrics; otherwise, the performance of the method is not evaluated correctly.

4. *What is the research gap to enhance HAR in P+L? What does the future road map look like?*

From the reviewer's perspective, further research on HAR for P+L should focus on five issues. First, a high-quality benchmark dataset for HAR methods to deploy in P+L is missing. This dataset should contain motion patterns that are as close to reality as possible and it should allow for comparison among different methods and thus being relevant for application in industry. Second, it must become possible to quantify the data creation effort, including both recording and annotation following a predefined protocol. This allows for a holistic effort estimation when deploying HAR in P+L. Third, most of the observed activities in the literature corpus are simplistic and they do not cover the entirety of manually performed work in P+L. Furthermore, the definition of activities cannot be considered fixed at design time and expected to remain the same during run time in such a rapidly evolving industry. Methods of HAR for P+L must address this issue. Fourth, method-wise, the segmentation approach should be revised in detail as a window-based approach is currently the only method for generating activity hypothesis.

This method does not handle activities that differ on their duration. A new method for computing activities with strongly different duration is needed. Fifth, the methods using deep networks do not include confidence measure. Even though these network methods show the state-of-the-art performance on benchmark datasets, they are still overconfident with their predictions. For this reason, integrating deep architectures with probabilistic reasoning for solving HAR using context information can be difficult.

Looking towards the future, the authors plan to tackle the issues along the outlined road map with further research.

**Author Contributions:** conceptualization, C.R. and F.N.; methodology, C.R. and F.N.; investigation, C.R., F.N. and F.M.R.; data curation, C.R., F.N. and F.M.R.; writing—original draft preparation, C.R., F.N. and F.M.R.; writing—review and editing, C.R., G.A.F. and M.t.H.; visualization, C.R.; supervision, C.R., G.A.F. and M.t.H.; project administration, G.A.F. and M.t.H.; funding acquisition, G.A.F. and M.t.H.

**Funding:** The work on this publication was supported by Deutsche Forschungsgemeinschaft (DFG) in the context of the research project Fi799/10-1, HO2403/14-1 (“Adaptive, Context-based Activity Recognition and Motion Classification to Analyze the Manual Order Picking Process”).

**Conflicts of Interest:** The authors declare no conflict of interest

## References

1. Dregger, J.; Niehaus, J.; Ittermann, P.; Hirsch-Kreinsen, H.; ten Hompel, M. Challenges for the future of industrial labor in manufacturing and logistics using the example of order picking systems. *Procedia CIRP* **2018**, *67*, 140–143, doi:10.1016/j.procir.2017.12.190. [[CrossRef](#)]
2. Hofmann, E.; Rüscher, M. Industry 4.0 and the current status as well as future prospects on logistics. *Comput. Ind.* **2017**, *89*, 23–34, doi:10.1016/j.compind.2017.04.002. [[CrossRef](#)]
3. Michel, R. *2016 Warehouse/DC Operations Survey: Ready to Confront Complexity*; Northwestern University Transportation Library: Evanston, IL, USA, 2016.
4. Schlögl, D.; Zsifkovits, H. Manuelle Kommissioniersysteme und die Rolle des Menschen. *BHM Berg-und Hüttenmänn. Monatshefte* **2016**, *161*, 225–228, doi:10.1007/s00501-016-0481-7. [[CrossRef](#)]
5. Liang, C.; Chee, K.J.; Zou, Y.; Zhu, H.; Causo, A.; Vidas, S.; Teng, T.; Chen, I.M.; Low, K.H.; Cheah, C.C. Automated Robot Picking System for E-Commerce Fulfillment Warehouse Application. In Proceedings of the 14th IFToMM World Congress, Taipei, Taiwan, 25–30 October 2015; pp. 398–403, doi:10.6567/IFToMM.14TH.WC.OS13.077. [[CrossRef](#)]
6. Oleari, F.; Magnani, M.; Ronzoni, D.; Sabbatini, L. Industrial AGVs: Toward a pervasive diffusion in modern factory warehouses. In Proceedings of the 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP), Piscataway, NJ, USA, 4–6 September 2014; pp. 233–238, doi:10.1109/ICCP.2014.6937002. [[CrossRef](#)]
7. Grosse, E.H.; Glock, C.H.; Neumann, W.P. Human Factors in Order Picking System Design: A Content Analysis. *IFAC-PapersOnLine* **2015**, *48*, 320–325, doi:10.1016/j.ifacol.2015.06.101. [[CrossRef](#)]
8. Calzavara, M.; Glock, C.H.; Grosse, E.H.; Persona, A.; Sgarbossa, F. Analysis of economic and ergonomic performance measures of different rack layouts in an order picking warehouse. *Comput. Ind. Eng.* **2017**, *111*, 527–536, doi:10.1016/j.cie.2016.07.001. [[CrossRef](#)]
9. Grosse, E.H.; Calzavara, M.; Glock, C.H.; Sgarbossa, F. Incorporating human factors into decision support models for production and logistics: Current state of research. *IFAC-PapersOnLine* **2017**, *50*, 6900–6905, doi:10.1016/j.ifacol.2017.08.1214. [[CrossRef](#)]
10. Chen, C.; Jafari, R.; Kehtarnavaz, N. A survey of depth and inertial sensor fusion for human action recognition. *Multimed. Tools Appl.* **2017**, *76*, 4405–4425, doi:10.1007/s11042-015-3177-1. [[CrossRef](#)]
11. Ordóñez, F.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115, doi:10.3390/s16010115. [[CrossRef](#)]
12. Haescher, M.; Matthies, D.J.; Srinivasan, K.; Bieber, G. Mobile Assisted Living: Smartwatch-based Fall Risk Assessment for Elderly People. In Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and Interaction iWOAR’18, Berlin, Germany, 20–21 September 2018; pp. 6:1–6:10, doi:10.1145/3266157.3266210. [[CrossRef](#)]

13. Hölzemann, A.; Van Laerhoven, K. Using Wrist-Worn Activity Recognition for Basketball Game Analysis. Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and Interaction iWOAR'18, Berlin, Germany, 20–21 September 2018; pp. 13:1–13:6, doi:10.1145/3266157.3266217. [[CrossRef](#)]
14. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, ICST, Austin, TX, USA, 6–7 November 2014; doi:10.4108/icst.mobibase.2014.257786. [[CrossRef](#)]
15. Feichtenhofer, C.; Pinz, A.; Zisserman, A. Convolutional Two-Stream Network Fusion for Video Action Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1933–1941, doi:10.1109/CVPR.2016.213. [[CrossRef](#)]
16. Ronao, C.A.; Cho, S.B. Deep Convolutional Neural Networks for Human Activity Recognition with Smartphone Sensors. In *Neural Information Processing; Lecture Notes in Computer Science*; Arik, S., Huang, T., Lai, W.K., Liu, Q., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 46–53, doi:10.1007/978-3-319-26561-2\_6.
17. Yang, J.B.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. In Proceedings of the 24th International Conference on Artificial Intelligence IJCAI'15, Buenos Aires, Argentina, 25–31 July 2015; pp. 3995–4001.
18. Bishop, C.M. *Pattern Recognition and Machine Learning*; Information Science and Statistics; Springer: Cham, Switzerland, 2006.
19. Fink, G.A. *Markov Models for Pattern Recognition: From Theory to Applications*, 2nd ed.; Advances in Computer Vision and Pattern Recognition; Springer: Cham, Switzerland, 2014.
20. Twomey, N.; Diethel, T.; Fafoutis, X.; Elsts, A.; McConville, R.; Flach, P.; Craddock, I. A Comprehensive Study of Activity Recognition Using Accelerometers. *Informatics* **2018**, *5*, 27, doi:10.3390/informatics5020027. [[CrossRef](#)]
21. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2015.
22. Yao, R.; Lin, G.; Shi, Q.; Ranasinghe, D.C. Efficient dense labelling of human activity sequences from wearables using fully convolutional networks. *y* **2018**, *78*, 252–266, doi:10.1016/j.patcog.2017.12.024. [[CrossRef](#)]
23. Feldhorst, S.; Aniol, S.; ten Hompel, M. Human Activity Recognition in der Kommissionierung—Charakterisierung des Kommissionierprozesses als Ausgangsbasis für die Methodenentwicklung. *Logist. J. Proc.* **2016** *2016*, doi:10.2195/lj\_Proc\_feldhorst\_de\_201610\_01. [[CrossRef](#)]
24. Alam, M.A.U.; Roy, N. Unseen Activity Recognitions: A Hierarchical Active Transfer Learning Approach. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA, 5–8 June 2017; pp. 436–446, doi:10.1109/ICDCS.2017.264. [[CrossRef](#)]
25. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359, doi:10.1109/TKDE.2009.191. [[CrossRef](#)]
26. Luan, P.G.; Tan, N.T.; Thinh, N.T. Estimation and Recognition of Motion Segmentation and Pose IMU-Based Human Motion Capture. In *Robot Intelligence Technology and Applications 5; Advances in Intelligent Systems and Computing*; Kim, J.H., Myung, H., Kim, J., Xu, W., Matson, E.T., Jung, J.W., Choi, H.L., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 383–391, doi:10.1007/978-3-319-78452-6\_32.
27. Pfister, A.; West, A.M.; Bronner, S.; Noah, J.A. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *J. Med. Eng. Technol.* **2014**, *38*, 274–280, doi:10.3109/03091902.2014.909540. [[CrossRef](#)] [[PubMed](#)]
28. Schlagenhauf, F.; Sahoo, P.P.; Singhose, W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. *Robot Autom. Eng. J.* **2017**, *1*, 555558, doi:10.19080/RAEJ.2017.01.555558. [[CrossRef](#)]
29. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 1–33, doi:10.1145/2499621. [[CrossRef](#)]
30. Roggen, D.; Förster, K.; Calatroni, A.; Tröster, G. The adARC pattern analysis architecture for adaptive human activity recognition systems. *J. Ambient. Intell. Humaniz. Comput.* **2013**, *4*, 169–186, doi:10.1007/s12652-011-0064-0. [[CrossRef](#)]

31. Dalmazzo, D.; Tassani, S.; Ramírez, R. A Machine Learning Approach to Violin Bow Technique Classification: a Comparison Between IMU and MOCAP systems Dalmazzo, David and Tassani, Simone and Ramírez, Rafael. In Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and InteractioniWOAR'18, Berlin, Germany, 20–21 September 2018; pp. 12:1–12:8, doi:10.1145/3266157.3266216. [[CrossRef](#)]
32. Vinciarelli, A.; Esposito, A.; André, E.; Bonin, F.; Chetouani, M.; Cohn, J.F.; Cristani, M.; Fuhrmann, F.; Gilmartin, E.; Hammal, Z.; et al. Open Challenges in Modelling, Analysis and Synthesis of Human Behaviour in Human–Human and Human–Machine Interactions. *Cogn. Comput.* **2015**, *7*, 397–413, doi:10.1007/s12559-015-9326-z. [[CrossRef](#)]
33. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209, doi:10.1109/SURV.2012.110112.00192. [[CrossRef](#)]
34. Xing, S.; Hanghang, T.; Ping, J. Activity recognition with smartphone sensors. *Tinshhua Sci. Technol.* **2014**, *19*, 235–249, doi:10.1109/TST.2014.6838194. [[CrossRef](#)]
35. Attal, F.; Mohammed, S.; Dedabrishvili, M.; Chamroukhi, F.; Oukhellou, L.; Amirat, Y. Physical Human Activity Recognition Using Wearable Sensors. *Sensors* **2015**, *15*, 31314–31338, doi:10.3390/s151229858. [[CrossRef](#)]
36. Edwards, M.; Deng, J.; Xie, X. From pose to activity: Surveying datasets and introducing CONVERSE. *Comput. Vis. Image Underst.* **2016**, *144*, 73–105, doi:10.1016/j.cviu.2015.10.010. [[CrossRef](#)]
37. O'Reilly, M.; Caulfield, B.; Ward, T.; Johnston, W.; Doherty, C. Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review. *Sport. Med.* **2018**, *48*, 1221–1246, doi:10.1007/s40279-018-0878-4. [[CrossRef](#)] [[PubMed](#)]
38. Kitchenham, B.; Brereton, P. A systematic review of systematic review process research in software engineering. *Inf. Softw. Technol.* **2013**, *55*, 2049–2075, doi:10.1016/j.infsof.2013.07.010. [[CrossRef](#)]
39. Kitchenham, B.; Pearl Brereton, O.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15, doi:10.1016/j.infsof.2008.09.009. [[CrossRef](#)]
40. Kitchenham, B. *Procedures for Performing Systematic Reviews*; Keele University: Keele, UK, 2004; p. 33.
41. Chen, L.; Zhao, X.; Tang, O.; Price, L.; Zhang, S.; Zhu, W. Supply chain collaboration for sustainability: A literature review and future research agenda. *Int. J. Prod. Econ.* **2017**, *194*, 73–87, doi:10.1016/j.ijpe.2017.04.005. [[CrossRef](#)]
42. Caspersen, C.J.; Powell, K.E.; Christenson, G.M. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep.* **1985**, *100*, 126–131. [[PubMed](#)]
43. Purkayastha, A.; Palmaro, E.; Falk-Krzyszinski, H.J.; Baas, J. Comparison of two article-level, field-independent citation metrics: Field-Weighted Citation Impact (FWCI) and Relative Citation Ratio (RCR). *J. Inf.* **2019**, *13*, 635–642, doi:10.1016/j.joi.2019.03.012. [[CrossRef](#)]
44. Xi, L.; Bin, Y.; Aarts, R. Single-accelerometer-based daily physical activity classification. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 6107–6110, doi:10.1109/IEMBS.2009.5334925. [[CrossRef](#)]
45. Altun, K.; Barshan, B. Human Activity Recognition Using Inertial/Magnetic Sensor Units. In *Human Behavior Understanding*; Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6219, pp. 38–51, doi:10.1007/978-3-642-14715-9\_5.
46. Altun, K.; Barshan, B.; Tunçel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.* **2010**, *43*, 3605–3620, doi:10.1016/j.patcog.2010.04.019. [[CrossRef](#)]
47. Khan, A.M.; Lee, Y.K.; Lee, S.Y.; Kim, T.S. Human Activity Recognition via an Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis. In Proceedings of the 2010 5th International Conference on Future Information Technology, Busan, Korea, 21–23 May 2010; pp. 1–6, doi:10.1109/FUTURETECH.2010.5482729. [[CrossRef](#)]
48. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explor. Newsl.* **2011**, *12*, 74, doi:10.1145/1964897.1964918. [[CrossRef](#)]
49. Wang, L.; Gu, T.; Tao, X.; Chen, H.; Lu, J. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive Mob. Comput.* **2011**, *7*, 287–298, doi:10.1016/j.pmcj.2010.11.008. [[CrossRef](#)]

50. Casale, P.; Pujol, O.; Radeva, P. Human Activity Recognition from Accelerometer Data Using a Wearable Device. In *Pattern Recognition and Image Analysis*; Vitrià, J., Sanches, J.M., Hernández, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6669, pp. 289–296, doi:10.1007/978-3-642-21257-4\_36.
51. Gu, T.; Wang, L.; Wu, Z.; Tao, X.; Lu, J. A Pattern Mining Approach to Sensor-Based Human Activity Recognition. *IEEE Trans. Knowl. Data Eng.* **2010**, *23*, 1359–1372, doi:10.1109/TKDE.2010.184. [[CrossRef](#)]
52. Lee, Y.S.; Cho, S.B. Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer. In *Hybrid Artificial Intelligent Systems*; Corchado, E., Kurzyński, M., Woźniak, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6678, pp. 460–467, doi:10.1007/978-3-642-21219-2\_58.
53. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. In *Ambient Assisted Living and Home Care*; Bravo, J., Hervás, R., Rodríguez, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7657, pp. 216–223, doi:10.1007/978-3-642-35395-6\_30.
54. Deng, L.; Leung, H.; Gu, N.; Yang, Y. *Generalized Model-Based Human Motion Recognition with Body Partition Index Maps*; Blackwell Publishing Ltd.: Oxford, UK, 2012; Volume 31, pp. 202–215, doi:10.1111/j.1467-8659.2011.02095.x.
55. Lara, S.D.; Labrador, M.A. A mobile platform for real-time human activity recognition. In Proceedings of the 2012 IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 14–17 January 2012; pp. 667–671, doi:10.1109/CCNC.2012.6181018. [[CrossRef](#)]
56. Lara, O.D.; Pérez, A.J.; Labrador, M.A.; Posada, J.D. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive Mob. Comput.* **2012**, *8*, 717–729, doi:10.1016/j.pmcj.2011.06.004. [[CrossRef](#)]
57. Siirtola, P.; Röning, J. Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data. *IJIMAI* **2012**, *1*, 38, doi:10.9781/ijimai.2012.155. [[CrossRef](#)]
58. Koskimäki, H.; Huikari, V.; Siirtola, P.; Röning, J. Behavior modeling in industrial assembly lines using a wrist-worn inertial measurement unit. *J. Ambient. Intell. Humaniz. Comput.* **2013**, *4*, 187–194, doi:10.1007/s12652-011-0061-3. [[CrossRef](#)]
59. Shoaib, M.; Scholten, H.; Havinga, P. Towards Physical Activity Recognition Using Smartphone Sensors. In Proceedings of the 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, Vietri sul Mare, Italy, 18–21 December 2013; pp. 80–87, doi:10.1109/UIC-ATC.2013.43. [[CrossRef](#)]
60. Zhang, M.; Sawchuk, A.A. Human Daily Activity Recognition With Sparse Representation Using Wearable Sensors. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 553–560, doi:10.1109/JBHI.2013.2253613. [[CrossRef](#)] [[PubMed](#)]
61. Bayat, A.; Pomplun, M.; Tran, D.A. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. *Procedia Comput. Sci.* **2014**, *34*, 450–457, doi:10.1016/j.procs.2014.07.009. [[CrossRef](#)]
62. Garcia-Ceja, E.; Brena, R.; Carrasco-Jimenez, J.; Garrido, L. Long-Term Activity Recognition from Wristwatch Accelerometer Data. *Sensors* **2014**, *14*, 22500–22524, doi:10.3390/s141222500. [[CrossRef](#)] [[PubMed](#)]
63. Gupta, P.; Dallas, T. Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1780–1786, doi:10.1109/TBME.2014.2307069. [[CrossRef](#)]
64. Kwon, Y.; Kang, K.; Bae, C. Unsupervised learning for human activity recognition using smartphone sensors. *Expert Syst. Appl.* **2014**, *41*, 6067–6074, doi:10.1016/j.eswa.2014.04.037. [[CrossRef](#)]
65. Aly, H.; Ismail, M.A. ubiMonitor: intelligent fusion of body-worn sensors for real-time human activity recognition. In Proceedings of the 30th Annual ACM Symposium on Applied Computing-SAC'15, Salamanca, Spain, 13–17 April 2015; pp. 563–568, doi:10.1145/2695664.2695912. [[CrossRef](#)]
66. Bleser, G.; Steffen, D.; Reiss, A.; Weber, M.; Hendebey, G.; Fradet, L. Personalized Physical Activity Monitoring Using Wearable Sensors. In *Smart Health*; Holzinger, A., Röcker, C., Ziefle, M., Eds.; Springer International Publishing: Cham, Switzerland, 2015; Volume 8700, pp. 99–124, doi:10.1007/978-3-319-16226-3\_5.
67. Chen, Y.; Xue, Y. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, China, 9–12 October 2015; pp. 1488–1492, doi:10.1109/SMC.2015.263. [[CrossRef](#)]
68. Guo, M.; Wang, Z. A feature extraction method for human action recognition using body-worn inertial sensors. In Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative

- Work in Design (CSCWD), Calabria, Italy, 6–8 May 2015; pp. 576–581, doi:10.1109/CSCWD.2015.7231022. [[CrossRef](#)]
69. Zainudin, M.; Sulaiman, M.N.; Mustapha, N.; Perumal, T. Activity recognition based on accelerometer sensor using combinational classifiers. In Proceedings of the 2015 IEEE Conference on Open Systems (ICOS), Bandar Melaka, Malaysia, 24–26 August 2015; pp. 68–73, doi:10.1109/ICOS.2015.7377280. [[CrossRef](#)]
  70. Ayachi, F.S.; Nguyen, H.P.; Lavigne-Pelletier, C.; Goubault, E.; Boissy, P.; Duval, C. Wavelet-based algorithm for auto-detection of daily living activities of older adults captured by multiple inertial measurement units (IMUs). *Physiol. Meas.* **2016**, *37*, 442–461, doi:10.1088/0967-3334/37/3/442. [[CrossRef](#)]
  71. Fallmann, S.; Kropf, J. Human Activity Pattern Recognition based on Continuous Data from a Body Worn Sensor placed on the Hand Wrist using Hidden Markov Models. *Simul. Notes Eur.* **2016**, *26*, 9–16, doi:10.11128/sne.26.tn.10322. [[CrossRef](#)]
  72. Feldhorst, S.; Masoudenijad, M.; ten Hompel, M.; Fink, G.A. Motion Classification for Analyzing the Order Picking Process using Mobile Sensors-General Concepts, Case Studies and Empirical Evaluation: In Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods, Rome, Italy, 24–26 February 2016; SCITEPRESS-Science and Technology Publications: Setubal, Portugal, 2016; pp. 706–713, doi:10.5220/0005828407060713. [[CrossRef](#)]
  73. Hammerla, N.Y.; Halloran, S.; Ploetz, T. Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. *arXiv* **2016**, arXiv:1604.08880.
  74. Liu, Y.; Nie, L.; Liu, L.; Rosenblum, D.S. From action to activity: Sensor-based activity recognition. *Neurocomputing* **2016**, *181*, 108–115, doi:10.1016/j.neucom.2015.08.096. [[CrossRef](#)]
  75. Margarito, J.; Helaloui, R.; Bianchi, A.; Sartor, F.; Bonomi, A. User-Independent Recognition of Sports Activities from a Single Wrist-worn Accelerometer: A Template Matching Based Approach. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 788–796, doi:10.1109/TBME.2015.2471094. [[CrossRef](#)] [[PubMed](#)]
  76. Reyes-Ortiz, J.L.; Oneto, L.; Samà, A.; Parra, X.; Anguita, D. Transition-Aware Human Activity Recognition Using Smartphones. *Neurocomputing* **2016**, *171*, 754–767, doi:10.1016/j.neucom.2015.07.085. [[CrossRef](#)]
  77. Ronao, C.A.; Cho, S.B. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **2016**, *59*, 235–244, doi:10.1016/j.eswa.2016.04.032. [[CrossRef](#)]
  78. Ronao, C.A.; Cho, S.B. Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models. *Int. J. Distrib. Sens. Netw.* **2017**, *13*, 155014771668368, doi:10.1177/1550147716683687. [[CrossRef](#)]
  79. Song-Mi, L.; Sangm M,Y.; Heeryon, C. Human activity recognition from accelerometer data using Convolutional Neural Network. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, Korea, 13–16 February 2017; pp. 131–134, doi:10.1109/BIGCOMP.2017.7881728. [[CrossRef](#)]
  80. Scheurer, S.; Tedesco, S.; Brown, K.N.; O’Flynn, B. Human activity recognition for emergency first responders via body-worn inertial sensors. In Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Eindhoven, The Netherlands, 9–12 May 2017; pp. 5–8, doi:10.1109/BSN.2017.7935994. [[CrossRef](#)]
  81. Vital, J.P.M.; Faria, D.R.; Dias, G.; Couceiro, M.S.; Coutinho, F.; Ferreira, N.M.F. Combining discriminative spatiotemporal features for daily life activity recognition using wearable motion sensing suit. *Pattern Anal. Appl.* **2017**, *20*, 1179–1194, doi:10.1007/s10044-016-0558-7. [[CrossRef](#)]
  82. Chen, Z.; Le, Z.; Cao, Z.; Guo, J. Distilling the Knowledge From Handcrafted Features for Human Activity Recognition. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4334–4342, doi:10.1109/TII.2018.2789925. [[CrossRef](#)]
  83. Moya Rueda, F.; Grzeszick, R.; Fink, G.; Feldhorst, S.; ten Hompel, M. Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors. *Informatics* **2018**, *5*, 26, doi:10.3390/informatics5020026. [[CrossRef](#)]
  84. Nair, N.; Thomas, C.; Jayagopi, D.B. Human Activity Recognition Using Temporal Convolutional Network. In Proceedings of the 5th international Workshop on Sensor-Based Activity Recognition and Interaction-iWOAR’18, Berlin, Germany, 20–21 September 2018; pp. 1–8, doi:10.1145/3266157.3266221. [[CrossRef](#)]
  85. Reining, C.; Schlangen, M.; Hissmann, L.; ten Hompel, M.; Moya, F.; Fink, G.A. Attribute Representation for Human Activity Recognition of Manual Order Picking Activities. In Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction-iWOAR’18, Berlin, Germany, 20–21 September 2018; pp. 1–10, doi:10.1145/3266157.3266214. [[CrossRef](#)]

86. Tao, W.; Lai, Z.H.; Leu, M.C.; Yin, Z. Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks. *Procedia Manuf.* **2018**, *26*, 1159–1166, doi:10.1016/j.promfg.2018.07.152. [[CrossRef](#)]
87. Wolff, J.P.; Grützmacher, F.; Wellnitz, A.; Haubelt, C. Activity Recognition using Head Worn Inertial Sensors. In Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction-iWOAR'18, Berlin, Germany, 20–21 September 2018; pp. 1–7, doi:10.1145/3266157.3266218. [[CrossRef](#)]
88. Xi, R.; Li, M.; Hou, M.; Fu, M.; Qu, H.; Liu, D.; Haruna, C.R. Deep Dilation on Multimodality Time Series for Human Activity Recognition. *IEEE Access* **2018**, *6*, 53381–53396, doi:10.1109/ACCESS.2018.2870841. [[CrossRef](#)]
89. Xie, L.; Tian, J.; Ding, G.; Zhao, Q. Human activity recognition method based on inertial sensor and barometer. In Proceedings of the 2018 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), Moltrasio, Italy, 26–29 March 2018; pp. 1–4, doi:10.1109/ISISS.2018.8358140. [[CrossRef](#)]
90. Zhao, J.; Obonyo, E. Towards a Data-Driven Approach to Injury Prevention in Construction. In *Advanced Computing Strategies for Engineering*; Smith, I.F.C., Domer, B., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 10863, pp. 385–411, doi:10.1007/978-3-319-91635-4\_20.
91. Zhu, Q.; Chen, Z.; Yeng, C.S. A Novel Semi-supervised Deep Learning Method for Human Activity Recognition. *IEEE Trans. Ind. Inform.* **2018**, 3821–3830, doi:10.1109/TII.2018.2889315. [[CrossRef](#)]
92. Rueda, F.M.; Fink, G.A. Learning Attribute Representation for Human Activity Recognition. *arXiv* **2018**, arXiv:1802.00761.
93. Lampert, C.H.; Nickisch, H.; Harmeling, S. Attribute-Based Classification for Zero-Shot Visual Object Categorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 453–465, doi:10.1109/TPAMI.2013.140. [[CrossRef](#)] [[PubMed](#)]
94. Lockhart, J.W.; Weiss, G.M.; Xue, J.C.; Gallagher, S.T.; Grosner, A.B.; Pulickal, T.T. *WISDM Lab: Dataset*; Department of Computer & Information Science, Fordham University: Bronx, NY, USA, 2013.
95. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. *WISDM Lab: Dataset*; Department of Computer & Information Science, Fordham University: Bronx, NY, USA, 2012.
96. Roggen, D.; Plotnik, M.; Hausdorff, J. *UCI Machine Learning Repository: Daphnet Freezing of Gait Data Set*; School of Information and Computer Science, University of California: Irvine, CA, USA, 2013. Available online: <https://archive.ics.uci.edu/ml/datasets/Daphnet+Freezing+of+Gait> (accessed on 20 July 2019).
97. Müller, M.; Röder, T.; Eberhardt, B.; Weber, A. *Motion Database HDM05*; Technical Report; Universität Bonn: Bonn, Germany, 2007.
98. Banos, O.; Toth, M.A.; Amft, O. UCI Machine Learning Repository: REALDISP Activity Recognition Dataset Data Set. Available online: <https://archive.ics.uci.edu/ml/datasets/REALDISP+Activity+Recognition+Dataset> (accessed on 20 July 2019).
99. Reyes-Ortiz, J.L.; Anguita, D.; Oneto, L.; Parra, X. UCI Machine Learning Repository: Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set. Available online: <https://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions> (accessed on 20 July 2019).
100. Zhang, M.; Sawchuk, A.A. Human Activities Dataset. 2012. Available online: <http://sipi.usc.edu/had/> (accessed on 20 July 2019).
101. Yang, A.Y.; Giani, A.; Giannantonio, R.; Gilani, K.; Iyengar, S.; Kuryloski, P.; Seto, E.; Seppa, V.P.; Wang, C.; Shia, V.; et al. d-WAR: Distributed Wearable Action Recognition. Available online: <https://people.eecs.berkeley.edu/~yang/software/WAR/> (accessed on 20 July 2019).
102. Roggen, D.; Calatroni, A.; Long-Van, N.D.; Chavarriaga, R.; Hesam, S.; Tejaswi Digumarti, S. UCI Machine Learning Repository: OPPORTUNITY Activity Recognition Data Set. Available online: <https://archive.ics.uci.edu/ml/datasets/opportunity+activity+recognition> (accessed on 20 July 2019).
103. Reyes-Ortiz, J.L.; Anguita, D.; Ghio, A.; Oneto, L.; Parra, X. UCI Machine Learning Repository: Human Activity Recognition Using Smartphones Data Set. Available online: <https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones> (accessed on 20 July 2019).
104. Reiss, A. UCI Machine Learning Repository: PAMAP2 Physical Activity Monitoring Data Set. Available online: <https://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring> (accessed on 20 July 2019).

105. Bulling, A.; Blanke, U.; Schiele, B. MATLAB Human Activity Recognition Toolbox. Available online: <https://github.com/andreas-bulling/ActRecTut> (accessed on 20 July 2019).
106. Zappi, P.; Lombriser, C.; Stiefmeier, T.; Farella, E.; Roggen, D.; Benini, L.; Tröster, G. Activity Recognition from On-body Sensors: Accuracy-power Trade-off by Dynamic Sensor Selection. In Proceedings of the 5th European Conference on Wireless Sensor Networks EWSN'08, Bologna, Italy, 30 January–1 February 2008; Springer: Berlin, Heidelberg, 2008; pp. 17–33.
107. Fukushima, K.; Miyake, S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognit.* **1982**, *15*, 455–469. [[CrossRef](#)]
108. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Curran Associates, Inc.: Red Hook, NY, USA; pp. 1097–1105.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).