Received August 9, 2018, accepted September 22, 2018, date of publication October 2, 2018, date of current version October 31, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2873502*

# Sensor-Based Datasets for Human Activity Recognition – A Systematic Review of Literature

**EMIRO DE-LA-HOZ-FRANCO[1], PAOLA ARIZA-COLPAS[1], JAVIER MEDINA QUERO[2], AND MACARENA ESPINILLA[2]**

[1]Department of Computer Science and Electronics, Universidad de la Costa–CUC, Barranquilla 080002, Colombia
[2]Department of Computer Science, University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain

Corresponding author: Emiro De-la-Hoz-Franco (edelahoz@cuc.edu.co)

This work was supported in part by the REMIND Project through the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie under Grant 734355.

**ABSTRACT** The research area of ambient assisted living has led to the development of activity recognition systems (ARS) based on human activity recognition (HAR). These systems improve the quality of life and the health care of the elderly and dependent people. However, before making them available to end users, it is necessary to evaluate their performance in recognizing activities of daily living, using data set benchmarks in experimental scenarios. For that reason, the scientific community has developed and provided a huge amount of data sets for HAR. Therefore, identifying which ones to use in the evaluation process and which techniques are the most appropriate for prediction of HAR in a specific context is not a trivial task and is key to further progress in this area of research. This work presents a systematic review of the literature of the sensor-based data sets used to evaluate ARS. On the one hand, an analysis of different variables taken from indexed publications related to this field was performed. The sources of information are journals, proceedings, and books located in specialized databases. The analyzed variables characterize publications by year, database, type, quartile, country of origin, and destination, using scientometrics, which allowed identification of the data set most used by researchers. On the other hand, the descriptive and functional variables were analyzed for each of the identified data sets: occupation, annotation, approach, segmentation, representation, feature selection, balancing and addition of instances, and classifier used for recognition. This paper provides an analysis of the sensor-based data sets used in HAR to date, identifying the most appropriate dataset to evaluate ARS and the classification techniques that generate better results.

**INDEX TERMS** Ambient assisted living–AAL, human activity recognition–HAR, activities of daily living–ADL, activity recognition systems–ARS, dataset.

## I. INTRODUCTION

The care of elderly dependent people who have difficulties to effectively develop ADL requires a lot of attention and dedication, because both the lifestyle and the health state of these people are affected. The proliferation of problems associated with dementia in older adults between 74 and 84 years of age [1] constitutes one of the main public health challenges worldwide. Due to this fact, secondary problems are generated that affect mental, physical and mobility capabilities [2]–[4]. In addition, there is a decline in basic communication skills, such as writing, speaking and performing simple and complex motor activities (cooking, taking medications and paying bills, among others) [5].

Nowadays, there has been a growing need for society to take care of their health integrating the use of technology. HAR enables monitoring of people's quality of life and more features and functionalities arise in this area over time, relying on a wide repertoire of hardware and software components. The research area of AAL has influenced the generation of reminder solutions, as a support for people suffering from neurodegenerative diseases. Proof of this is the implementation of several solutions in indoor environments, which capture the data generated from the interactions of people with an intelligent environment [6]. The objectives of HAR, based on the analysis of ADL [7], are: 1) the creation of predictive models that allow the classification of the normal and abnormal behaviour of individuals [8], 2) to provide the necessary tools for the caregiver and the medical team to identify the activities carried out by them and generate preventive and corrective measures.

The data collected from heterogeneous sensors deployed in smart environments or from sensors attached to the body (wearables), are stored in datasets. In this way, different modalities of data collection have been proposed: video [9], [10], audio [11], [12] and binary sensors [6], [13] or portable sensors deployed on the body such as accelerometers and gyroscopes [14], [15], among others. The dataset is then used to train different machine learning techniques that predict the behaviour of people with different purposes, such as sending early warnings to caregivers and mitigating the risks related to the deterioration of the health of the monitored people.

Currently there is a large amount of datasets for HAR. Therefore, identifying which ones to use in the evaluation process of an ARS and which techniques (in the phases of preprocessing, extraction features, feature selection and transformation, classification and post-classification) are the most appropriate to improve the rates of Activity Recognition (AR) is a complex task. The exploratory process of identifying the most appropriate dataset to be used in the evaluation of the ARS and the identification of the techniques that have been successfully applied in different AR approaches, in order to improve accuracy rates, demands considerable time that the researcher could use on other tasks at the core of their research.

ARS have emerged thanks to the advancement of sensor technology, mainly for its ability to understand the situations that arise in contexts in which humans interact while performing ADL. As indicated in [16], the practical applications of ARS are numerous: fall detection [17], gait anomaly detection [18], energy expenditure estimation [19], [20], stress detection [21], behaviour monitoring [22] and rehabilitation [23], among others. Therefore, evaluating the reliability of ARS in terms of their ability to predict the different activities collected in the dataset is a challenging task.

Initially, ARS were evaluated with adapted laboratory datasets, which were recorded in controlled conditions. With the growing development of new ARS, the difficulty of collecting data is increased, since the collection of data recorded in a particular laboratory does not include a wide enough variety of activities to evaluate ARS with sufficient rigor. Given this situation, a series of dataset benchmarks (Opportunity [24], HASC [25], AmI Repository [26], among others) have emerged. Additionally, the scientific community has created a competition called Evaluating AAL Systems Through Competitive Benchmarking – AR (EvAAL-AR) [27]. Both initiatives aim for researchers: 1) to develop ARS and put them to the test in the context of experimentation, 2) to submit their proposals to evaluation using different dataset benchmarks, and 3) to validate their developments in an academic competition.

Motivated by this research field, the main contribution of this paper is:

1) The identification of the most recognised datasets by the academic community regarding HAR, assessing the types of activity and data, data capture devices, level of occupation, annotation, context and scenario where the data have been collected, the duration of the capture and the number of individuals or inhabitants that generated the activities.

2) A characterisation and analysis of each identified dataset, which includes: the different classification techniques used for the AR, the segmentation techniques used to select the data-streaming windows, the feature representation of the data, the distribution of the dataset used for training and testing and the quality metrics (F-measure and accuracy) used in the HAR evaluation, and

3) A series of suggestions in relation to the use of different techniques: segmentation, feature representation, feature selection, balancing or addition of instances and distribution of datasets for the experimentation processes.

In previous literature, two studies [28], [29] can be identified, in which six dataset benchmarks are compared to several ARS. In the first study, a comparative analysis of six benchmark datasets (Berkeley [30], USC [31], HMD [32], Opportunity [24], UTD MHAD [33] and Sport [34]) was made comparing the F-measure obtained in nine ARS proposals, including theirs (identifying the classifier or classifiers used in each proposal). In the second study, a comparative analysis of 15 proposals (including theirs) was documented, comparing whether they were used or not: six dataset benchmarks (CASAS [23], VanKasteren [6] and others), two feature selection techniques (Principal Component Analysis – PCA and Information Gain – IG) and 12 association approaches (or classifiers). The proposed approach in this paper is an original contribution, because there is no benchmark that has identified the most relevant datasets in terms of HAR. In this paper, for each one of them, a detailed characterisation and subsequent analysis is made in relation to the different classification techniques used for AR including the segmentation techniques used to select the data-streaming windows, the feature representation of the data, the distribution of the dataset used for training and testing, and the quality metrics (F-measure and accuracy) used in the HAR evaluation.

The compilation, documentation and analysis of the aforementioned variables for each dataset and the studies done between 2003 and 2017 constitute a considerable volume of papers to be reviewed. This task of evaluating the particularities of each variable analysed in the datasets using scientific rigor makes this work a valuable and relevant contribution.

The work has been structured in the following way: in Section II, related research is described. Section III proposes the methodology used for the systematic review of the literature. In Section IV, the scientometric analysis is presented and, in Section V, the technical analyses are presented. Section VI presents the characterisation of each of the identified datasets and the respective analysis of results.

Section VII presents the discussion and, finally, Sections VIII and IX present the conclusions and future work, respectively.

## II. RELATED RESEARCH

When consulting related works, it has been observed that some reviews of the literature analyse very specific approaches from the point of view of technology (Mode of Movement Recognition – MMR [35], Acceleration-Based Activity – ABC [36] and Non-Intrusive Load Monitoring NILM [37]). Other reviews, albeit not thoroughly, focus on the evaluation of classification techniques in terms of their effectiveness for the recognition of activities of daily life from datasets [38], [39]. However, there is no evidence that proposes a systematic review of the literature, which allows characterising datasets for HAR based on the analysis of ADL. Below a detailed description of the studies previously mentioned is provided.

In [35] the approaches to MMR are compared and described from different viewpoints: usability and convenience, device types, data collection methods, types and errors of sensors used, signal pre-processing methods employed (windowing, de-noising and variable calculation), feature extraction (statistical and time-domain features, energy, power, magnitude and frequency-domain), feature selection and transformation, classification techniques and post-classification refining. This paper ends with a quantitative comparison of the performance of motion mode recognition modules developed by researchers in different domains.

In [36], a naturalistic 3D acceleration-based activity dataset, the SCUT-NAA dataset (publicly available) is created to assist researchers in the field of acceleration-based AR and to provide a standard dataset for comparing and evaluating the performance of different algorithms. The SCUT-NAA contains 1278 samples from 44 subjects, collected in naturalistic settings with only one tri-axial accelerometer located alternatively on the waist belt, in the trouser pocket, and in the shirt pocket. In this research they showed a summary of some representative datasets (none publicly available) on AR using acceleration, in which they compared: number of activities, number of subjects, whether data was collected under laboratory or naturalistic settings, number of accelerometers used per subject and accuracy.

In [37], a novel technique to monitor human activity based on NILM is presented. In order to evaluate the performance of the proposed algorithm, two different datasets have been considered: the Household Electricity Survey dataset [40] and the UK Domestic Appliance-Level Electricity – UK-DALE dataset [41]. Both use real collected data from the aggregated and disaggregated energy consumption of UK households. The former contains a year of data from three single pensioner households, which are the targeted community in this study; whereas the latter is a two-year collection of data from a family household (two adults, two children and a dog).

In [38], they proposed the use of Deep Learning (DL) techniques to automatically learn high-level features from binary sensor data. To evaluate the performance of the proposed method, they applied experiments to three (publicly available) smart home datasets, and compared it with a range of shallow models in terms of time-slice accuracy and class accuracy. The datasets were collected using simple sensors (motion detector, contact switch, pressure mats, mercury contacts and float sensors), and each of the smart homes housed one resident performing ADL. The description of each of the datasets, in relation to the number of sensors, data collection time, number of activities, resulting sensor events and activity instances is as follows: first dataset (14 sensors, 25 days, 10 activities, 1229 sensor events and 292 activity instances), second dataset (23 sensors, 14 days, 13 activities, 19075 sensor events and 200 activity instances) and third dataset (21 sensors, 19 days, 16 activities, 22700 sensor events and 344 activity instances).

In [39], three learning classification algorithms were implemented to evaluate the AR of ADL using: Naïve Bayesian (NB), Support Vector Machine (SVM) and Random Forest (RF). For this, recruiting ten healthy subjects and monitoring their activities over 20 days using the sensor system was necessary.

## III. METHODOLOGY

The SRL is a key piece of secondary research that allows the creation of frameworks on which future research is supported. An outstanding reference in this respect is [42], which proposes a methodology based on the definition of research questions, search process, inclusion and exclusion criteria, quality assessment, data collection, data analysis and deviations from protocol. The vast majority of research of this type is carried out in the health field, for which the methodologies proposed in [43]–[45] were analysed. In the field of engineering, [42] and [46] propose their respective methodologies. Reference [46] is organised in three stages: the definition of search parameters (objective, hypothesis and search index), identification and debugging in data bases (selection of search chains whose results will be deepened) and the proposal of answers to the hypothesis (from the information obtained from the categorisation and analysis of the most relevant articles). Specifically, in [47] a methodology was proposed to perform a validation of Chronic Kidney Disease (CKD) and related conditions in existing datasets (including administrative datasets and disease registries). All these studies contributed to the approach of the SRL methodology used in the research described here, which was organised in three stages.

The definition of search parameters configures the first stage and consists in determining the objectives of the review, to then identify the following hypothesis: "Which datasets have provided the greatest impact on the development of research related to the recognition of ADL?" Subsequent to this, we proceeded to locate the topics of the databases on which the search would focus (Scopus, IEEExplore, Science Direct, Web of Science and ACM). From these, the keywords to be used and discarded were identified and validated, due to the noise generated by the latter in the results. Fig. 1 show

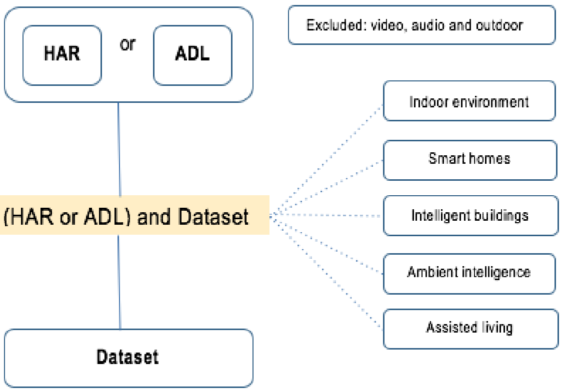


**FIGURE 1.** Search chains conformation model.

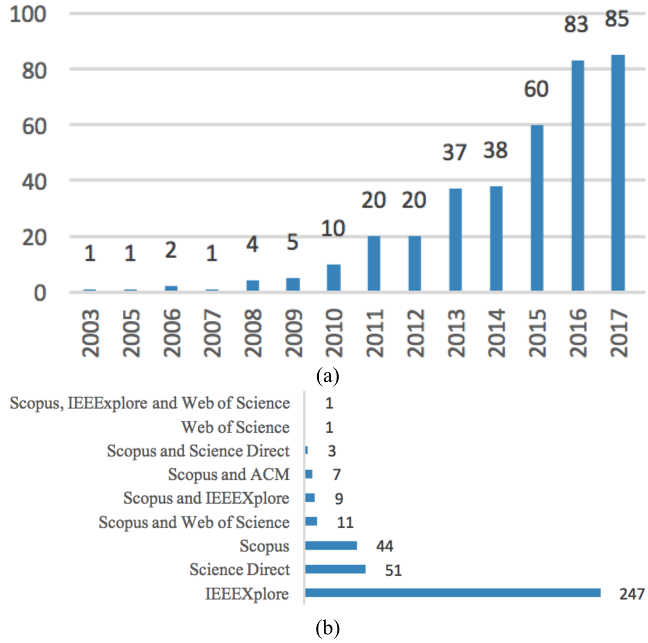


**FIGURE 2.** Trend in the number of publications: (a) number of publications by year and (b) publications by scientific database.

the schema from which the search terms were built, where the term "video" and "audio" have been excluded, given that the processing of data based on this type of research differs considerably from other feedback mechanisms such as pressure, contact, positioning and accelerometer sensors, among others. Additionally, the term "outdoor" was excluded in order to delimit the scope of the publications to be analysed.

In the second stage, the identification and filtering of the information obtained from the specialised databases was undertaken, and the results obtained from the application of the different search chains were analysed. The data have been represented synthetically in different arrays and where the search indexes are hierarchically organised, discarding the combinations that did not yield any results. In addition, the terms that yielded a large number of results were identified, with greater specificity criteria added to limit the searches to the subject matter of the research.

The search chains that yielded results were selected, identifying the articles that match the proposed hypothesis. The search chains were constructed using the keywords identified in Fig. 1, its structure being as follows: (HAR OR ADL) AND dataset AND ("indoor environment" OR "smart homes" OR "intelligent buildings" OR "ambient intelligence" OR "assisted living") AND NOT (video) AND NOT (audio) AND NOT (outdoor). A series of scientometric variables were documented for each article, such as the year of publication, the journal, the typology of the document, the journal's quartile, the country of publication of the journal, the country from where the production is generated and the entity or university that presents the product. Additionally, a series of technical variables were documented to characterise the type of dataset used in the different research works consulted. Such variables were: 1) if reference was made to a dataset or a repository, 2) its name, 3) what type of events it contained, 4) its level of occupation and annotation, 5) the sensing modality with which the dataset was fed. Finally, the approach, segmentation, representation, feature selection, balancing and addition of instances, category and subcategory of the classifiers referenced in the papers. The last stage is the presentation and analysis of results.

## IV. SCIENTOMETRIC ANALYSIS

After recording the scientometric variables of the 374 publications, they were quantified based on the following criteria: year of publication, number of articles published by database (identifying those that were referenced in several databases), publications according to the typology of the same and to the quartile of the magazine, congress or book where it was published. We also considered the identification of the countries that receive a greater flow of works and those that output a greater flow of works, the journals and universities that have more development in this concrete field of research. The following figures and tables of contents illustrate the above with greater precision.

Evidence of the validity of this field of research (see Fig. 2) is the growing trend in the number of publications related to HAR, in terms of: 1) the implementation of intelligent environments in indoor contexts, 2) the capture of data generated from the interactions of the inhabitants with the sensors deployed in such environments, 3) the collection and structuring of datasets and 4) the application of predictive algorithms for the classification of ADL. Fig. 2a shows how the number of publications has grown each year in this research field reaching its peak in 2017. Fig. 2b indicates that between 2003 and 2017, the scientific database with the most research products registered (Journals, Proceedings and Chapters of books, among others) is IEEEXplore.

66% of the publications in this field of knowledge are carried out in journals and proceedings, which can be accessed from the IEEExplore specialised database. 72% of the publications accessible from the different specialised databases are carried out in proceedings and 27% in journals (see Fig. 3a). Although it is true that the highest percentage of

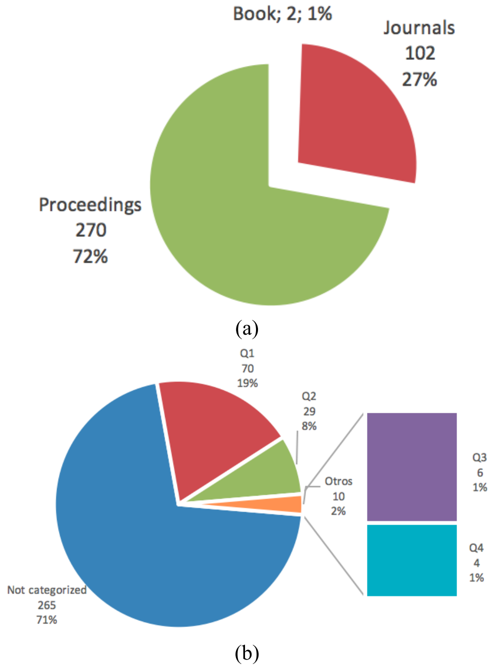


**FIGURE 3.** Publications by type and quartile: (a) number of publications by type and (b) publications by quartile.

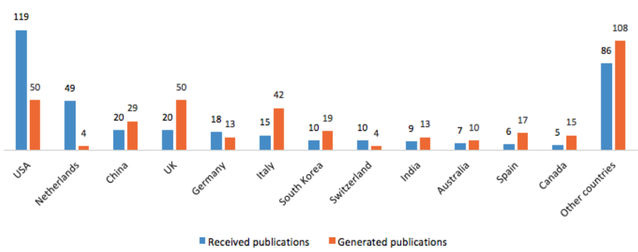


**FIGURE 4.** Number of publications received and generated by country.

publications is made in uncategorised resources (71%), there is a significant percentage (19%) of publications that are made in journals of the first quartile (see Fig. 3b).

Fig. 4 shows the number of publications received and generated by country, between 2003 and 2017. The first value was calculated by counting the publications by the country of origin or edition of the journal, proceeding or book. From this, it has been identified that the countries with the highest number of publications received, in relation to the scope of the HAR, are: USA, Netherlands, China, UK and Germany, among others. To account for the publications by the country of generation of these, the place of origin of the university, research center or organization to which the authors of said publication are affiliated was identified. Specifically, the criterion of identification of the country of generation, was determined mostly by taking the one that was most common to all the authors of the respective paper and in some exceptional cases, taking the place of origin of the first author. From this, we have identified that the countries with the highest number of publications generated, based on HAR, are: USA, UK, Italy, China and South Korea, among others.

**TABLE 1.** Journals with the most publications.

| | Publications | Country | Type | ISSN | Quartile |
|---|---|---|---|---|---|
| Pervasive and Mobile Computing | 10 | Netherlands | Journal | 15741192 | Q1 |
| Lecture Notes in Computer Science | 10 | Germany | Book Series | 03029743 | Q2 |
| Procedia Computer Science | 9 | Netherlands | Proceeding | 1877-0509 | --- |
| IEEE Journal of Biomed. and Health Informatics | 8 | USA | Journal | 21682208 21682194 | Q1 |
| IEEE Sensors Journal | 5 | USA | Journal | 1530-437X | Q1 |
| Neurocomputing | 5 | Netherlands | Journal | 0925-2312 | Q1 |
| Expert Systems with Applications | 4 | UK | Journal | 9574174 | Q1 |
| Sensors | 4 | Switzerland | Journal | 1424-8220 | Q2 |

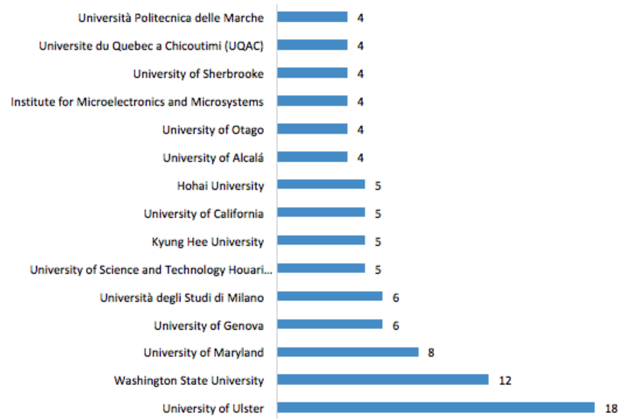


**FIGURE 5.** Publications by University.

Table 1 shows the resources where more works related to the recognition of ADL are published. It is noteworthy that most of them present a good ranking in relation to their quartile.

In addition, we note the production in research at the level of scientific Journals regarding pervasive healthcare, by the University of Ulster and Washington State University – WSU (see Fig. 5). This fact is due to the institutions having generated and validated their own datasets, which have been widely used by a representative sector of the academic community. In this way, we highlight the University of Ulster, which presented an initiative for the creation of open datasets within pervasive healthcare, which can be consulted in [48]. For its part, WSU is the creator of the most complete repository of AR in smart homes, called CASAS [23], [49].

## V. TECHNICAL ANALYSIS

The Inter-University Consortium for Political and Social Research – ICPSR [50] is an international consortium of more than 750 academic institutions and research organisations, which provides leadership and training in data access, curation, and methods of analysis for the social science research community. This organisation has Institutional Review Boards (IRB) which review research proposals [51]. It is a good practice to submit ADL recognition datasets to

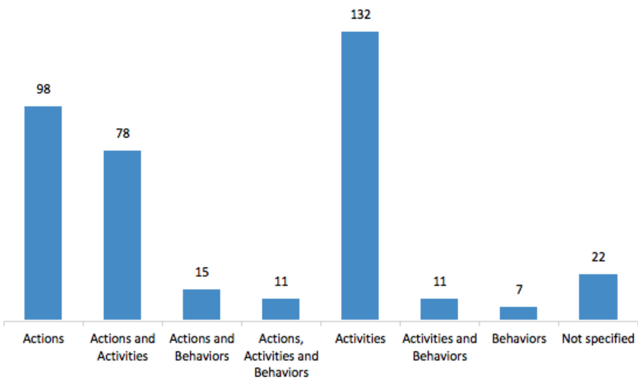


**FIGURE 6.** Documented events in the datasets.

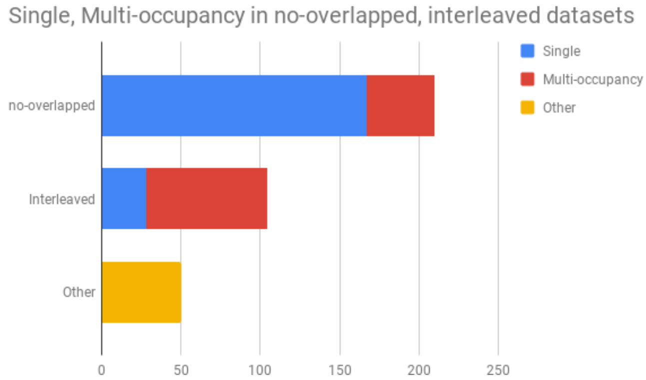


**FIGURE 7.** Level of occupations.

the IRB for review, in order to safeguard the human subjects who participate in biomedical or behavioural research.

Rodríguez *et al.* [52] classify the dataset according to the type of event or activity granularity levels they record: actions, activities and behaviours, First, the actions or atomic events with a timestamp define the lowest granularity degree of representation (e.g.: open door, move object, turn light off, walk by, be observed in location, among others). Second, the activities, which were considered as single actions with an inherent purpose or composed by a set of different actions, represent an intermediate level of representation as regards granularity and have a start-date-time and end-date-time (e.g.: take coffee, attend conference, group meeting, video call, send email, among others). Last, the behaviours were defined as a sequence of activities and/or actions by a set of compulsory actions or activities plus a set of optional actions or activities, where some of them can have temporal execution interdependencies (e.g., the behaviour coffee break includes the action exit office, the activity make coffee or take coffee, and the action enter office in this order). Fig. 6 shows the number of papers that reference actions, activities or behaviours (or combinations of these events).

According to [53], we identify many datasets proposed in the literature in which three classes of activities can be distinguished: single activity [54], interleaved activity [55] and multi-occupancy [13]. Single defines an activity which has been fully carried out before starting the performance of a new one; interleaved activities are carried out while another activity is being performed at the same time; and multi-occupancy is related to a class of activity in which some people are performing their activities simultaneously. Fig. 7 indicates that, in relation to occupation, the vast majority of the papers consulted have used datasets that contain data records of single activities, 42.2% of the papers reviewed, while 20.3% correspond to datasets with interleaved activities.

Otherwise, a dataset is annotated when each data record is assigned a class tag that identifies it. In this case, researchers use different techniques (manual and automatic) to assign the labelling of the dataset records. Fig. 8 shows that the majority of datasets referenced in the papers are annotated (60.4%).

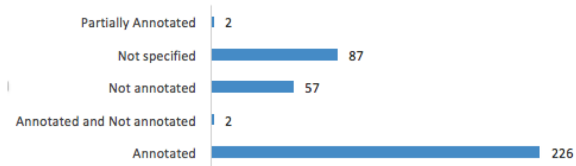


**FIGURE 8.** Number of papers that relate the type of dataset annotation.

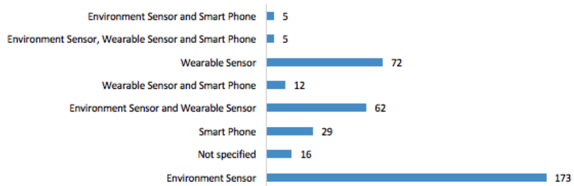


**FIGURE 9.** Number of papers according to sensing modalities.

In Fig. 9, we show the results according to the modality in which the data were sensed to feed the dataset. It is noteworthy that most of the research has been done where the data is captured by the use of environmental sensors (46.3%), although there is an evident growth in wearables sensors (19.3%).

A relevant classification of Machine Learning is presented in [48], which shows two approaches for AR: Data-Driven Approaches (DDA) and Knowledge-Driven Approaches (KDA). DDA are based on machine learning techniques in which a pre-existent dataset of user behaviours is required. Here a training process is usually carried out to build an activity model which is followed by a testing process and to evaluate the generalisation of the model in classifying unseen activities [38], [108]. Regarding KDA, an activity model is built through the incorporation of rich prior knowledge gleaned from the application domain, using knowledge engineering and knowledge management techniques [109], [110]. The vast majority of studies reviewed (76.2%), mention the name of the classifier used (285 papers), while the remaining 23.8% do not mention it (89 papers), 60.9% use classifiers based on DDA, 12.6% use classifiers based on KDA and 2.7% use classifiers of both approaches.

**TABLE 2. References to classifier subcategories.**

| Approach | Category | Subcategory | References | % |
|---|---|---|---|---|
| DDA | MLM | MM | 156 | 41.7% |
| | | SVM | 65 | 17.4% |
| | | IBL | 59 | 15.8% |
| | | BC | 59 | 15.8% |
| | | DT | 36 | 9.6% |
| | | ANN | 28 | 7.5% |
| | | Clustering | 7 | 1.9% |
| KDA | MLC | MaS | 28 | 7.5% |
| | | Cascading | 5 | 1.3% |
| | SABL | Semantic models | 6 | 1.6% |

**TABLE 3. Most referenced dataset.**

| Dataset | Event | Datatype | Ref. | % | Devices |
|---|---|---|---|---|---|
| VanKasteren [57] | Activities | Binary values | 20 | 11% | Wireless Sensor Network – WSN and Sensors |
| CASAS Kyoto [23] | Activities | Date, Time, SensorID, Value (bin. or num.) | 14 | 8% | Motion, associated with objects and telephone sensors. |
| CASAS Aruba [58] | Activities | Date, Time, SensorID, Value (bin. or num.) | 14 | 8% | Motion, door and temperature sensors. |
| CASAS Multiresident [13] | Activities | Date, Time, SensorID, Value (bin.), ResidentID and TaskID | 12 | 7% | Motion, item, cabinet, water, burner, phone and temperature sensors. |
| UCI HAR [59-60] | Actions | Normalised to values between 1 and -1 | 10 | 6% | Accelerometer and gyroscope. |
| Opportunity [24], [61-62] | Activities | Text file (array, each row is a sample) | 8 | 5% | Inertial sensors and Accelerometers. |
| mHealth [63], [80] | Actions | Fine-grained real-valued sensor readings of actions over a short time interval | 6 | 3% | Accelerometer, ECG, Gyroscope and Magnetometer. |

**TABLE 4. Details of the most referenced dataset.**

| Dataset | Occupancy | Context | Sensing | Capture | Individuals |
|---|---|---|---|---|---|
| Van Kasteren [57] | Single | Two houses (kitchen and bathroom). | Environment Sensor | Two weeks | Two men |
| CASAS Kyoto [23] | Single | WSU smart workplace. | Environment Sensor | Not specified | 20 participants |
| CASAS Aruba [58] | Multi-occupancy | Washington State University smart workplace. | Environment Sensor | Not specified | An older volunteer woman, and her children and grandchildren. |
| CASAS Multiresident [13] | Multi-occupancy | Washington State University smart apartment. | Environment Sensor | One month | Two participants |
| UCI HAR [59-60] | N/A | Handset mounted: on the left-side of the belt and placed according to the user's preference. | Smart Phone | Not specified | 30 volunteer subjects. |
| Opportunity [24], [61-62] | Interleaved and hierarchical naturalistic activities. | A room simulating a studio flat | Environment and on-body sensors. | Not specified | 4 subjects. |
| mHealth [63], [80] | N/A | Chest, wrist and ankle. | On-body sensors. | Not specified | 10 subjects. |

In this study, to account for the results obtained, the hierarchy defined in [35] was used, in relation to the approaches, categories, subcategories and classification techniques, used for HAR based on the identification of ADL. In several papers, reference is made to more than one classification technique. Therefore, in some cases, the reference to several subcategories of techniques was counted for the same papers. Table 2 contains the number of references within these subcategories associated with the respective categories and approaches (the 89 papers that do not mention the name of the classifier used were discarded). The most referenced subcategories were those belonging to the DDA categories of Machine Learning Methods – MLM (e.g., Markov Models – MM, Instance Based Classifiers or Instance based learning – IBL, Bayesian Classifiers – BC, Decision Trees – DT, Artificial Neural Networks – ANN, among others); the subcategories belonging to Meta-Level Classifiers – MLC (e.g., Multi-agent System – MaS and Cascading) and Semantic Attribute-Based Learning (SABL) within KDA categories are referenced to a lesser extent.

The consulted papers refer to one or more classification techniques, with the purpose of carrying out a comparative analysis of their performance. Accordingly, 41.7% of the references to classifiers correspond to the MM subcategory (specifically using the Hidden-Markov Model – HMM classifier), while 17.4% reference the SVM classifier, and the subcategories IBL and BC have 15.8% references each in the papers consulted, the first one specifically using the k-Nearest Neighbor – kNN algorithm and the second using NB and BC classifiers (see Table 2).

On the other hand, of 374 publications related to this field of research, 94% of the works (352 papers) indicate that they have used datasets (their own or from other authors), while 2% (7 papers) do not specify what type of data structure has been used (if it is a dataset or a repository) and the remaining 4% (15 papers) mention the use of a repository, in which the results obtained from the processing are compared to the different datasets that constitute it. The most used repositories are CASAS [23], [49], [54] and UCI Machine Learning Repository [56]. It is important to note that some papers mention one or several datasets generated or used. Of the 352 papers that indicate the use of datasets, 50% (175 papers) explicitly mention the name of the dataset used. In some of these papers, reference is made to more than one dataset when a comparative analysis of the performance quality metrics of the classifier used is performed. The remaining 50% (177 papers) do not mention the name of the dataset used. Tables 3 and 4 show the details of the seven (7) most referenced datasets (all are annotated), in the 175 publications that do identify the name of the dataset used.

The Van Kasteren dataset [57] is the result of the measurement of a Wireless Sensor Network (WSN) in an enclosure that is occupied by two men (26 and 57 years old). In this apartment there are 14 sensors that indicate changes of state associated with actions such as: opening and closing of doors, pressure on the apartment floor, as well as sensors on the bed and on the sofa. The characteristics of the dataset are: the stored values are binary (either because of the use of binary sensors for the capture or because some threshold was applied to the captured analogue value), it has 245 actions from different activities (brushing teeth, showering, toileting, bathing, shaving, breakfast, dinner, snacking, drinking, loading the dishwasher, unloading the dishwasher, among others) and the duration of the capture process was two weeks (which included sensor data and annotation). The data were captured through the implementation of RFID, WSN and different types of sensors (reed switches, mercury contacts, passive infrared - PIR and float sensors). For the annotation process, a combination of Bluetooth headsets with speech recognition and a handwritten register of activities were used. In [2], [29], [64], and [65] different classification techniques (NB, HMM, Hidden Semi-Markov Model – HSMM and Conditional Random Field – CRF) were used to compare Van Kasteren with one or more datasets.

The CASAS project [23], [49], [54] is located on the campus of WSU. The apartment is made up of a bathroom, a living room, three bedrooms and a kitchen. The sensors in the apartment are distributed at a distance of approximately one meter. The sensors can be categorised as: motion sensor, motion area sensor (covers a larger region), item sensor for selected items in the kitchen, door sensor, burner sensor, hot water sensor, cold water sensor, temperature sensor, electricity usage, battery level, light level, shake sensor, light sensor, gyro sensor, experimenter switch (manual trigger) and fan. The official website[1] of the project contains a wide variety of datasets and tools. For each dataset the following is detailed: the name of the testbed, the number of residents or participants, whether it is annotated or not. Additionally, the files of each dataset are available for download. The CASAS repository also has the following tools: real-time activity profiling, activity learning (recognition, discovery, and prediction), AR, rule-based activity prediction, pattern visualiser, activity visualisation, real-time annotation tools, data sampling tools, sequential prediction, multi-view transfer learning techniques and mobile activity learner (IOS and Android).

The most referenced datasets of the CASAS repository are: Daily life 2010-2012 (Testbed: Kyoto) also called ADL Activities [23], Daily life 2010-2011-2012 (Testbed: Aruba) [58] and Multiresident ADL Activities (Testbed: Kyoto) [13]. The first contains activities (making a call, washing hands, cooking, eating and washing the dishes) collected from 20 participants, using environment sensors (motion, associated with objects, from the medicine box, a flowerpot, a diary, a closet, water, kitchen and telephone use sensors), the dataset contains information related to the date and time of each event, the sensor ID and value (binary or numeric) of each sensor activated during the event. The second contains activities (movement from bed to bathroom, eating, getting home, housework, leaving home, preparing food, relaxing, sleeping, washing dishes and working) collected from an older volunteer woman, and her children and grandchildren (who visited her several times). They interacted with environment sensors (motion, door and temperature sensors) and, like the previous dataset, this one contains information related to the date and time of each event, the sensor ID and value (binary or numeric) of each sensor activated during the event. The third dataset contains a wide variety of activities (filling medication dispenser, hanging up clothes, moving the couch and coffee table, sitting on the couch, watering plants, sweeping the kitchen floor, playing a game of checkers, setting out ingredients for dinner, setting dining room table, reading a magazine, simulating the payment of an electric bill, gathering food for a picnic, retrieving dishes from a kitchen cabinet, packing supplies in the picnic basket and packing food in the picnic basket) collected from two inhabitants, who participated at the same time and with 26 tests for each pair of inhabitants (40 participants in total), where they interacted with environment sensors (motion, item, cabinet, water, burner, phone and temperature sensors). This dataset contains information related to the date and time of each event, the sensor ID and (binary) value of each sensor activated during the event and the Task ID identifies that event. The annotation process was manual (labelled during the recording). In [29], [64], and [66]–[73], different classification techniques were used to compare one of the datasets available in the CASAS repository, with one or more datasets of the same or other repositories. In these works, the most referenced CASAS datasets are: ARAS, Cairo, Aruba, Tulum, Kyoto (ADL Activities and Multiresident ADL Activities), DOMUS and Tokyo.

The HAR using a smartphone dataset – HAR [59], [60] was collected using a Samsung Galaxy SII smartphone, with the collaboration of 30 volunteers in order to identify actions (walking, going upstairs, going downstairs, sitting, standing and laying down). This dataset is available in the UCI Machine Learning dataset repository. For each record in the dataset, the following is provided: triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration, triaxial angular velocity from the gyroscope, to 561-feature vector with time and frequency domain variables, activity label and an identifier of the subject who carried out the experiment. In [74]–[79], different classification techniques were used to compare HAR datasets with one or more datasets of the same or other repositories.

The Opportunity AR dataset [24] (also available in the UCI repository), allows multi-patient experimentation on 4 subjects in the same venue. The subjects were monitored for 30 days, in a room simulating a studio flat with kitchen, deckchair, and outdoor access where subjects performed daily morning activities. The dataset contains activities (waking up, grooming, making breakfast and cleaning). The data capture is achieved through 68 sensors (14 located on objects, 21 ambient sensors and 33 located on the body). The data is provided as a text file containing an array where each row corresponds to a sample, the first column includes the sample timestamp (ms), and the last two columns include the labels for modes of locomotion and gestures respectively.

[1] http://casas.wsu.edu/

The annotation process was labelled while recording. In [16], [28], and [76] different classification techniques were used to compare Opportunity with one or more datasets.

The mHealth dataset [63], [80] (also available in the UCI repository), contains body motion and vital sign recordings for ten volunteers performing several physical actions (standing still, sitting and relaxing, lying down, walking, climbing stairs, bending waist forward, front arm elevation, knee bending, cycling, jogging, running, jumping front and back). Sensors were located on the chest, right wrist and left ankle of the subject and were used to measure the motion experienced in diverse body parts (acceleration, rate of turn and magnetic field orientation). The sensor positioned on the chest also provides 2-lead ECG measurements, which can be potentially used for basic heart monitoring, checking for various arrhythmias or looking at the effects of exercise on the ECG. The dataset includes fine-grained real-valued sensor readings of actions over a short time interval, with no explicit timestamps or locations included in the dataset. The annotation process was recorded using a video camera. In [75] and [81], [82] different classification techniques were used to compare mHealth with one or more datasets.

## VI. CHARACTERISATION AND ANALYSIS OF RESULTS

Regarding the classification techniques used for the processing of the aforementioned datasets, Tables 5 and 6 detail each of them according to the different approaches (DDA and KDA). Additionally, the total number of citations of the papers that introduce each dataset (taken from Google Scholar) are indicated. Some of the techniques referenced in the papers were not listed in the table, because there is no evidence that they have been applied to these datasets. However, the following are mentioned: Activity Discovery (AD) [83], Bayesian Belief Network (BBN) [84], Hierarchical, Autonomic Recursive and Distributed Bayesian Network (HARD-BN) [85], Cross-subject unsupervised transfer learning (CsUTL) [86], Data-Driven Non-Linear Hebbian (DD-NHL) [87], Dynamic Background Subtraction (DBS) [88], Temporal Learning using Echo State Network (TL-ESN) [89], Expectation Maximization (EM) algorithm [90], Extended Episode Discovery (xED) algorithm [91], Finite Action-set Learning Automata (FALA) [92], Finite State Machine (FSM) [93], Fuzzy Logic (FL) [94], Fuzzy HMM (FHMM) [95], Fuzzy Inference Model (FIM) [96], Fuzzy Temporal Relationships (FTR) [97], Learning Frequent Patterns of User Behaviour System (LFPUBS) [8], Minimum Redundancy Maximum Relevance (MRMR) [98], Multi-stage Decision Model (MsDM) [99], Qualitative Spatial Reasoning + AtomGID (QSR-AtGID) [100], Self-Adaptive Neural Networks (SANN) and Growing Self Organizing Maps (GSOM) [101], Semantic Indoor Trajectory Model and N-gram Model (SITM-NgM) [102], Sequential Extreme Learning Algorithm (SELA) [103], Suitability of Multi-label Learning Algorithms (SMLLA) [104], Term Based Labelling (TBL) [105] and User Behaviour Shift Detection (UBSD) [106].

**TABLE 5. KDA and classification techniques used in analysed datasets.**

| Dataset | Citation | KDA | Technique |
|---|---|---|---|
| Van Kasteren [57] | 840 | MLC | AR approach by Clustering based Classification - AR-CbC [107] |
| CASAS Kyoto [23] | 217 | MLC | HMM [117] |
| CASAS Aruba [58] | 236 | MLC | AR-CbC with Evidence Theoretic kNN - ET-kNN [107] |
| | | | Dynamic Bayesian Networks – DBN and HMM [120] |
| CASAS Multiresident [13] | 167 | - | - |
| UCI HAR [59-60] | 301 | MLC | Multiple HMM with Mixture-of-Templates - MOT and k-NN [78] |
| | | | HMM + Dictionary Learning - DicL [123] |
| Opportunity [24], [61-62] | 404 | MLC | Three-Stage Classifier - 3SC [125] |
| | | | Expert Rules to Recognize Postures - ER2RP + Data Mining Algorithms - DMA [16] |
| | | | Multilayer Perceptron - MLP + SVM + BN [126] |
| | | | DL (Non-negative Matrix Factorization - NMF + Stacked Auto-Encoder - SAE) [28] |
| mHealth [63], [80] | 103 | MLC | Hierarchical Classification Model - HCM (two base binary classifiers - 2bbC + Selective Learning of Slacked Hierarchy - SLoSH + Unified Objective Function - UOF) [81] |

A comprehensive review of each of the seven aforementioned datasets was carried out. With the analyses presented below, we aim to identify: the different classification techniques used for AR, the segmentation techniques used to select the data-streaming windows, the feature representation of the data, the distribution of the dataset used for training and testing, and the quality metrics (F-measure and accuracy) used in the evaluation of the proposals analysed. Not all studies applied the same quality metrics and not all present the values obtained with their corresponding standard deviation. The papers that do not explicitly indicate the techniques and metrics used for the processing and evaluation of the dataset were not documented in the respective tables.

### A. VANKASTEREN DATASET ANALYSIS

Table 7 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to the VanKasteren. From this, the following analysis was made:

- Whenever the VanKasteren and ARAS datasets were compared when [64] using Streaming Multi-Class imbalance ensemble NB classifiers (streamingMEn) and when [113] using HMM and HHMM, the best results were obtained with VanKasteren.
- Only in [38] the noise was eliminated, applying the Stacked Denoising Autoencoder (SDAE) technique, which generated an improvement in accuracy, when compared with not using this technique.
- In [29], [107], and [38] feature selection techniques were applied to improve the performance of the classifier. In the first one, an accuracy of up to 85.6% was obtained using the SAE technique. In the second, the IG technique was used, obtaining an accuracy of 95.3%. This significant result is also due to the fact that in this proposal, data balancing was used through the oversampling approach,

**TABLE 6. DDA and classification techniques used in analysed datasets.**

| Dataset | DDA | Technique |
|---|---|---|
| Van Kasteren [57] | MLM | ET-kNN [29], [107] |
| | | NB [2], [38], [64-65], [108-110] |
| | | Active Learning considering Overlapped activities - AALO [108] |
| | | HMM [2], [38], [57], [65], [68], [108-113] |
| | | HSMM [2], [38], [57], [108-109], [112] |
| | | Hierarchical HMM - HHMM [113] |
| | | DL [38] |
| | | SVM [2], [68], [38], [111], [114] |
| | | 1NN [38] |
| | | kNN [111] |
| | | Principal Component Classification and Correlation Criterion - PCC + CC [38], [115] |
| | | Dynamic Graphical Model – DGM [109] |
| | | Frequent Item set Mining - FrISM [110] |
| | | CRF [2], [57], [65], [68], [109], [111], [114] |
| | | Recurrent Neural Network – RNN [2] |
| | | Neighbourhood Counting Matrix - NCM [116] |
| | | ANN [68] |
| CASAS Kyoto [23] | MLM | Linear-Chain CRF – LCCRF [117] |
| | | Skip Chain CRF – SCCRF [117] |
| | | Difference of Convex Programming – DCP [117] |
| | | ET-kNN [29] |
| CASAS Aruba [58] | MLM | Multinomial Logistic Regression – MLR [118-119] |
| | | ET-kNN [107] |
| | | NB [67], [73] |
| | | HMM [73], [120] |
| | | CRF [73] |
| | | SVM [73], [121] |
| CASAS Multiresident [13] | MLM | NB [66] |
| | | HMM [66] |
| | | RF [122] |
| | | Viterbi [66] |
| UCI HAR [59-60] | MLM | Continuous HMM - CHMM [124] |
| | | SVM [78] |
| | | DT [78], NB [78] |
| | | MLP [78], RF [78] |
| | | Logistic Regression – LR [78] |
| | | HMM [78], [123] |
| Opportunity [24], [61-62] | MLM | Recursive Partitioning – RPART [125] |
| | | DT [16], [125], [127] |
| | | Bagging with DT - Bag-DT [125] |
| | | SVM [125-126] |
| | | NB [16], [125], [127] |
| | | Linear Discriminant Analysis – LDA [125] |
| | | RF [16], [125], [127] |
| | | kMeans Clustering – kMC [127] |
| | | Random Committee – RC [127] |
| | | Deep Canonical Correlation Analysis – dCCA [127] |
| | | MLP [126], BC [126] |
| mHealth [63], [80] | MLM | SVM [75], [81-82] |
| | | HMM [82] |
| | | Hidden CRF – HCRF [82] |
| | | Convolutional Neural NetWorks - CNN [82] |
| | | DT [81], kNN [81], AdaBoost [81] |
| | | Linear Discriminant Function – LDF [81] |
| | | Quadratic Discriminant Function - QDF [81] |

**TABLE 7. VanKasteren dataset evaluation.**

| Ref. | Segment. | Feature repr. | Activity Classification Technique | Folds | F-measure | Accuracy |
|---|---|---|---|---|---|---|
| [29] | AR-SPM | | IG + ET-KNN | | | **95.3** |
| [64] | | | streamingMEn | | 90.6 | |
| [38] | Tb-SW of 60 sec | Numeric | DL (two-layer: SDAE) | Lodo-CV | | 85.6 |
| | | | DL (one-layer: DAE) | | | 85.1 |
| | | | SVM | | | 81.8 |
| | | | 1NN | | | 59.0 |
| | | | HSMM | | | 69.4 |
| | | | HMM | | | 66.8 |
| | | | NB | | | 80.5 |
| **[115]** | | Lf | **PCC + CC** | Lodo-CV | | **96.9** |
| [109] | Explicit segment. (by activity) | Binary | DGM + NB | Lodo-CV | 50.5 | 78.4 |
| | | | DGM + HMM | | 50.2 | 70.0 |
| | | | DGM + HSMM | | 51.5 | 70.9 |
| | | | DGM + CRF | | 60.2 | 81.8 |
| [110] | Segment event stream | | NB, HMM and FIM | 80% training - 20% test | | 91.0 |
| [2] | Tb-SW of 60 sec | Lf | NB | Lodo-CV | | $95.3 \pm 2.8$ |
| | | | HMM | | | $89.5 \pm 8.4$ |
| | | | HSMM | | | $91.0 \pm 7.2$ |
| | | | CRF | | | $96.4 \pm 2.4$ |
| | | | Vanilla RNN | | | $95.5 \pm 3.4$ |
| | | | **LSTM - RNN** | | | $\mathbf{96.7 \pm 2.6}$ |
| | | | Gated Recurrent Unit RNN | | | $96.1 \pm 2.5$ |
| | | | SVM | | | $96.1 \pm 2.4$ |
| [114] | Tb-SW of 60 sec | Cp and Lf | C-SVM | Lodo-CV | | 92.1 |
| | | | CRFs | | | 92.0 |
| [111] | Tb-SW of 60 sec | Raw feature | SVM | Lodo-CV | | 85.6 |
| | | | HMM | | | 68.5 |
| | | | kNN | | | 73.1 |
| | | | CRFs | | | 89.8 |
| [65] | Tb-SW of 60 sec | Cp and Lf | NB | Lodo-CV | $65.0 \pm 1.6$ | |
| | | | CRF | | $70.0 \pm 1.6$ | |
| | | | HMM | | $72.0 \pm 1.5$ | |
| | | | 2-layer HMM | | $76.0 \pm 1.3$ | |
| [57] | Tb-SW of 60 sec | Cp and Lf | HMM | Lodo-CV | 57.3 | |
| | | | HSMM | | 76.4 | |
| | | | CRF | | 73.1 | |
| [113] | Tb-SW of 60 sec | | HMM | | 64.7 | |
| | | | HHMM | | 70.7 | |
| [116] | Sensor segment. | | NCM | | | 86.1 |
| [107] | Seg-AI | # times (sensor activ.) | AR-CbC with ET-kNN | Lodo-CV | | **96.3** |
| [68] | | | ANN | | | 41.1 |
| | | | HMM | | | 57.8 |
| | | | CRF | | | 92.3 |
| | | | SVM | | | 89.2 |

using the Synthetic Minority Oversampling Technique (SMOTE) [128]. Finally, in [107] PCA was used, obtaining an accuracy of 96.3%.

- Other studies that applied data balancing techniques were: [64] using Multi-class Stream Imbalance (McSI), [114] using an algorithmic approach and [2] adding manually synthesised abnormal activities (occurring at a wrong time of the day and after and before a specific activity); the latter is the proposal with the best accuracy: 96.7% $\pm$ 2.6.
- Although different segmentation techniques have been applied, the most commonly used in these studies was a Time-based and Sliding Window (Tb-SW) of 60 seconds. The best accuracy obtained when applying this segmentation technique was 96.7% in [2], using a classifier based on RNN, called Long Short Term (LSTM). However, it is important to note that a very high accuracy (96.9%) was obtained in [115] without using segmentation techniques, by applying the classification technique PCC + CC. Additionally, significant values of accuracy were obtained in [107] and [29] applying the Segmented Activity Instances (Seg-AI) (96.3%) and Activity Recognition - Segmental Pattern Mining (AR-SPM) (95.3%) techniques.
- The majority of the analysed studies use the Last-fired (Lf) and Change-point (Cp) feature representation techniques, where the highest accuracy was obtained when implemented in [115] 96.9% and [2] 96.7% $\pm$ 2.6.
- Leave one day out cross validation (Lodo-CV) was the most used distribution of data for training and testing in the different experimentation scenarios.
- The most commonly used classifiers in the analszed studies are those based on MM (in their different variations: HMM, HSMM and HHMM), NB, CRF and SVM,

individually employed, assembled with another technique or as a benchmark to compare with other proposed methods.

- The best accuracy achieved in AR, using the VanKasteren dataset, has been obtained in proposal [115] 96.9%, using PCC + CC as a classifier and in proposal [2] 96.7% ± 2.6, using the RNN called LSTM as a classifier.
- Proposals such as [108], where the metric evaluated was the Average time slice error (%), and [112] where the metrics evaluated were precision, recall and F1score, are not documented in Table 7, due to the uniformity of the analysis.

### B. CASAS-KYOTO DATASET ANALYSIS

Table 8 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to Kyoto. From this, the following analysis was made:

- In the analysed studies there is no evidence of the application of noise elimination techniques and feature representation when using Kyoto.
- There is only evidence of such in the application of two data segmentation techniques, in [73] and [117] Segments of Activities (Seg-oA) and in [29] AR-SPM, where best accuracy was obtained with the application of the latter.
- The distribution of data for training and testing in [29] and [117] was Leave one out cross validation (Loo-CV) with k-folds = 20 and in [73] with k-folds = 10.
- In [29], a comparison was made between VanKasteren and Kyoto, evaluating the performance of the ET-kNN, where best accuracy was obtained when applied in Kyoto (97.4%) using IG and overlapping activity classes as a feature selection technique.
- The best accuracy results using Kyoto were obtained in [29] at 97.4%, when the IG feature selection technique was applied and at 96.2% when it was not applied. This significant result is also due to the use of data balancing in this proposal through the over-sampling approach, using the SMOTE [128]. Although in [73] the combination of feature selection techniques Consistency Subset Eval (best First Search) – CsubE and Chisquared Attribute Evaluation (Ranker Thresh-old 1) – Ch2AE were applied, the obtained results did not exceed those previously mentioned.
- The most commonly used classifiers in the analysed studies are: LCCRF (with two variations according to the way the activities were evaluated: 1. Single Model for a Single Activity – SMSA and 2. Single Model for All Activities – SMAA), SCCRF, HMM, DCP, ET-kNN (applying and not IG) and SVM.

### C. CASAS-ARUBA DATASET ANALYSIS

Table 9 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to Aruba. From this, the following analysis was made:

**TABLE 8. CASAS-kyoto dataset evaluation.**

| Ref. | Segment. | Activity Classification Technique | Folds | F-measure | Accuracy |
|---|---|---|---|---|---|
| [117] | Seg-oA | LCCRF - SMSA | Loo-CV | 63.6 | 72.2 |
| | | LCCRF - SMAA | | 67.3 | 73.5 |
| | | SCCRF | | 48.6 | 60.1 |
| | | HMM | | 73.5 | 76.8 |
| | | DCP | | 82.3 | 82.6 |
| [29] | AR-SPM | **IG + ET-kNN** | Loo-CV | | **97.4** |
| | | **ET-kNN** | | | **96.2** |
| [73] | Seg-oA | SVM (Non-graph feat.) | 10-fold | | 83.2 |
| | | SVM (Graphical feat.) | | | 83.2 |

**TABLE 9. CASAS-aruba dataset evaluation.**

| Ref. | Segment. | Feature repr. | Activity Classification Technique | Folds | F-measure | Accuracy |
|---|---|---|---|---|---|---|
| [118] | D-win (# act sensors) | IG | MLR | 10- fold | 70.1 | |
| [119] | D-win (# act sensors) | IG | MLR | 10- fold | 87.0 | |
| [107] | Seg-oA | PCA | AR-CbC with ET-kNN | 3- fold | | 91.4 |
| [73] | Seg-oA | CsubE and Ch2AE | NB | 10-fold | | 91.2 |
| | | | HMM | | | 88.9 |
| | | | CRF | | | 91.8 |
| | | | SVM (Non-graph feat.) | | | 93.3 |
| | | | SVM (Graphical feat.) | | | 93.4 |
| [121] | Modified Sensor Win. Mutual Inf. Ext. | | SVM (Fixed-partition method - FIXED) | | | 53.9 |
| | | | SVM (Error-driven method - ERRD) | | | 49.6 |
| | | | SVM (Least Significant Support Vector Forget) | | | 46.5 |
| [120] | Win. of 30 sec | | DBN + HMM | 10-fold | | 80.5 |
| [67] | A-win | AFMSI-Act | **NB** | 80% train – 20% test | | **100.0** |

- In [73], a comparison was made between Aruba and Kyoto, evaluating the performance of the SVM classifier (both non-graphical and graphical features), where the best accuracy was obtained for Aruba at 93.4%, with graphical features.
- In the analysed studies, there is no evidence of the application of noise elimination techniques or addition of instances when using Aruba.
- Only two proposals use feature representation: [107] uses the number of times that sensors activated during activity and [121] uses Last-state (Ls) representation, the best accuracy between these two was obtained with the first at 91.4%.
- Different feature selection techniques were used: IG in [118] and [119], PCA in [107], CsubE and Ch2AE in [73] and Activity Features Maintain the Statistical Information about the Activities – AFMSI-Act (Mutual Information, Frequency of triggered sensors of an activity, Interval time and Last two sensors) in [67]. This last technique improved the accuracy of the classifier, achieving 100% in AR.
- Regarding the distribution of data for training and testing, although most of the studies used 10-fold

**TABLE 10. CASAS-multiresident dataset evaluation.**

| Ref. | Activity Classification Technique | Folds | F-measure | Accuracy |
|---|---|---|---|---|
| [66] | **NB** | 3-fold | | **81.5** |
| | HMM | | | 72.5 |
| | Viterbi | | | 73.5 |
| [122] | RF (multi-label classification) | 10 fold | 81.1 | 70.3 |

**TABLE 11. UCI-har dataset evaluation.**

| Ref. | Feature selection | Activity Classification Technique | Accuracy |
|---|---|---|---|
| [124] | RF (error rate = 0.08) | CHMM | 91.8 |
| | Stepwise LDA (error rate = 0.38) | | |
| | Correlation (error rate = 0.45) | | |
| | PCA (error rate = 0.54) | | |
| [78] | | SVM | 90.9 |
| | | DT (C4.5) | 80.4 |
| | | NB | 83.5 |
| | | MLP | 84.8 |
| | | RF | 87.6 |
| | | LR | 81.1 |
| | | HMM | 75.5 |
| | | **Multiple HMMs + MOT + k-NN** | **92.6** |
| [123] | PCA | HMM + DL | 90.5 |
| [75] | | SVM (Cs-RBFk) | 89.3 |
| | | SVM (Cs-Lk) | 90.8 |
| | | SVM (Cs-Sk) | 82.6 |
| | | SVM (Cu-mk) | 90.7 |
| | | SVM (Cu-mRBFk) | 89.6 |
| | | SVM (Cab-mk) | 91.3 |
| | | SVM () | 90.2 |
| | | **SVM (C-SNRb-mk)** | **91.4** |
| | | SVM (C-SNRb-mRBFk) | 90.9 |

cross-validation, the best result was obtained in [67] with a training of 80% and a test with 20% of the data.

- The analysed studies use different segmentation techniques: Dynamic Window (D-win) or Adaptive Window (A-win), which is based on the number of sensors activated or in the approach, by Segmentation of Activities (Seg-oA) and fixed size based on time. The highest 100% accuracy was obtained with an innovative segmentation technique called Adaptive windowing approach [67], which is divided into two phases (off-line modelling and on-line recognition). In the former phase a representation called Activity Features (AFs) is built from statistical information about the activities from annotated sensory data and a NB classifier is modelled accordingly. In the second phase, a dynamic multi-feature windowing approach using AFs and the learner NB classifier is introduced to segment unlabelled sensor data as well as predicting the related activity, as indicated in [67].
- The most used classifiers in the analysed studies are: MLR, ET-kNN, NB, HMM, CRF and SVM.

### D. CASAS-MULTIRESIDENT DATASET ANALYSIS

Table 10 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to the Multiresident ADL Activities (testbed: Kyoto) dataset. From this, the following analysis was made:

- In the analysed studies there is no evidence of the application of segmentation techniques, noise elimination and balancing or addition of instances when using the Multiresident ADL Activities dataset.
- Regarding the feature representation in [122], Change-point (Cp) was used, reaching an accuracy of 70.3%.
- In both [122] and [66] cross-validation was used, in the first with k-folds = 10 and in the second with k-folds = 3, the best accuracy being obtained in the latter.
- The best accuracy was 81.5% in [66], using the NB classifier and selecting features manually.

### E. UCI-HAR DATASET ANALYSIS

Table 11 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to UCI-HAR. From this, the following analysis was made:

- In the analysed studies there is no evidence of the application of noise elimination techniques and addition or balancing of instances when using UCI-HAR.
- Regarding the segmentation in [124], a sliding windows of 128 samples was established and in [123] a windows of raw signals is used, but no further details are given. In both studies feature representation techniques were used: in [124] they used Scaling to obtain z-scores (from values between 1 and -1), reaching an accuracy of 91.76% and in [123] they used raw signals representation, reaching 90.5%.
- The only study that explicitly indicates the distribution of the dataset in the experimentation process for training and testing is [124], where 10-fold cross validation was used, 70% for training (of 21 subjects) and 30% for testing (of 9 subjects), reaching an accuracy of 91.76%.
- In [75] different kernel approaches for the SVM classifier were analysed (Compressive single RBF kernel – Cs-RBFk, Compressive single Laplacian kernel – Cs-Lk, Compressive single sigmoid kernel – Cs-Sk, Compressive uniform multi-kernel – Cu-mk, Compressive uniform multi-RBF-kernel – Cu-mRBFk, Compressive alignment-based multi-kernel – Cab-mk, Compressive alignment-based multi-RBF-kernel – Cab-mRBFk, Compressive SNR-based multi-kernel – C-SNRb-mk and Compressive SNR-based multi-RBF-kernel – C-SNRb-mRBFk). When comparing their performance in the mHealth and UCI-HAR, the highest accuracy (91.4%) was obtained when applying SVM C-SNRb-mk in UCI-HAR.
- The best accuracy achieved with UCI-HAR was 92.6%, using Multiple HMMs classifier with MOT and kNN ensemble, proposed in [78], without applying feature selection techniques.
- The most used classifiers in the analysed studies are: SVM and HMM.

**TABLE 12. Opportunity dataset evaluation.**

| Ref. | Activity Classification Technique | Folds | Accuracy |
|---|---|---|---|
| [125] | RPART | 10-fold | 87.6 ± 1.4 |
| | DT | | 86.0 ± 1.5 |
| | Bag-DTs | | 89.4 ± 1.9 |
| | SVM | | 93.8 ± 1.2 |
| | NB | | 77.7 ± 3.3 |
| | LDA | | 89.3 ± 1.9 |
| | RF | | 93.1 ± 1.8 |
| | 3SC + RPART | | 87.5 ± 2.0 |
| | 3SC + DT | | 88.2 ± 2.4 |
| | 3SC + Bag-DTs | | 90.6 ± 1.7 |
| | **3SC + SVM** | | **94.0 ± 2.0** |
| | 3SC + NB | | 80.7 ± 4.3 |
| | 3SC + LDA | | 91.1 ± 1.2 |
| | 3SC + RF | | 93.9 ± 1.6 |
| [127] | kMC | 5-fold | 60.0 |
| | NB | | 79.0 |
| | DT | | 94.0 |
| | **RF** | | **96.3** |
| | RC | | 95.9 |
| | dCCA | | 96.2 |
| **[16]** | **ER2RP + DMA** | Leave-one-person-out cross-validation | **99.0** |
| | DT | | 96.4 |
| | NB | | 97.9 |
| | RF | | 98.3 |
| [126] | MLP | 10-fold random cross-validation | 92.5 ± 1.3 |
| | SVM | | 91.8 ± 1.4 |
| | BN | | 89.1 ± 1.3 |
| | **Fusion (MLP+SVM+BN)** | | **92.7 ± 1.3** |
| [28] | **DL (NMF+SAE)** | | **99.9** |

**TABLE 13. mHEALTH dataset evaluation.**

| Ref. | Activity Classification Technique | Accuracy |
|---|---|---|
| [75] | SVM (Cs-RBFk) | 78.3 |
| | SVM (Cs-Lk) | 78.1 |
| | SVM (Cu-mk) | 87.1 |
| | SVM (Cu-mRBFk) | 81.1 |
| | SVM (Cab-mk) | 79.4 |
| | SVM (Cab-mRBFk) | 84.1 |
| | **SVM (C-SNRb-mk)** | **87.3** |
| | SVM (C-SNRb-mRBFk) | 85.7 |
| [82] | HMM | 66.6 |
| | SVM | 65.4 |
| | HCRFs | 68.6 |
| | CNN with 1D kernel | 90.4 |
| | CNN with 2D kernel | 89.1 |
| | CNN - pf | 91.3 |
| | **CNN - pff** | **91.9** |
| [81] | DT | 67.0 |
| | LDF | 65.8 |
| | QDF | 74.9 |
| | SVM Linear | 55.2 |
| | SVM Gaussian | 52.1 |
| | 1-NN | 57.4 |
| | 5-NN | 57.0 |
| | AdaBoost | 68.2 |
| | QDF Hierarchical | 74.6 |
| | SVM Linear Hierarchical | 76.3 |
| | Ensemble (Hierarchical) | 86.1 |
| | HCM (2bbC + SLoSH + UOF) | **97.2** |

### *F. OPPORTUNITY DATASET ANALYSIS*

Table 12 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to Opportunity. From this, the following analysis was made:

- In the analysed studies there is no evidence of the application of noise elimination techniques when using Opportunity.
- The only study that explicitly claims to use segmentation techniques is [16] with a sliding window of 2 seconds (1 second overlap), reaching an accuracy of 99.0% when using a hybrid model based on ER2RP and a classifier trained with DMA.
- Regarding the feature representation in [125], the raw sensor data was compressed, which allowed to reach an accuracy of 94.0% ± 2.0, using a 3SC with SVM.
- Only two studies, [28] and [127], applied feature selection techniques: IG was used in the former, reaching an accuracy of 96.3% when applying the RF classifier, and in the latter, SAE were used, achieving an accuracy of 99.9% when applying a classifier based on DL Technique.
- Regarding the balancing or addition of instances, in [126] the instances are randomly selected for each fold among the four inhabitants, which allowed an accuracy of 92.7% ± 1.3 to be achieved, using a classifier that operated MLP, SVM and BC. A better result was achieved in [125] where the Instance Reassignment technique was applied (avoiding class imbalance within each group and also reducing the number of classes per group), which allowed to reach an accuracy of 94.0% ± 2.0 using a 3SC classifier with SVM.
- The best accuracy was 99.9% in [28], using a classifier based on DL technique. The approach proposed in [28] first applied to matrix factorization (NMF) in order to project data into a new reduced space to find a better activity representation and to increase discrimination capacity. Then, features were automatically extracted from the projected data using SAE. For classification, they built a softmax classifier on the top hidden layer of the SAE.
- In [127], Smartphone Activity Recognition (AR) [129] and Opportunity datasets were compared, while in [28] Berkeley [30], USC [31], HMD [32], UTD MHAD [33], Sport [34] and Opportunity datasets were compared, and in both studies Opportunity showed better results.
- The most used classifiers in the analysed studies are: SVM, RF, DT, NB and MLP.

### *G. mHEALT DATASET ANALYSIS*

Table 13 presents the evaluation of the pre-processing and classification techniques, as well as the quality metrics, of each of the proposals applied to mHealth. From this, the following analysis was made:

- In the analysed studies there is no evidence of the application of noise elimination techniques, features representation and selection techniques, balancing or addition of instances, nor the distribution of the datasets in the experimentation process (for training and testing) in mHealth.
- In [82] sliding window segmentation with a size of 60 samples was applied, obtaining an accuracy of 91.9% when using the CNN classification technique.

**TABLE 14. Classification techniques with better accuracy by dataset.**

| # | dataset | Activity Classification Technique | Accuracy |
|---|---|---|---|
| 1 | VanKasteren | Long Short Term RNNs [2] | 96.7 ± 2.6 |
| 2 | CASAS-Kyoto | IG + overlapping activity classes + ET-kNN [29] | 97.4 |
| 3 | CASAS-Aruba | NB classifier as a learner with Bernoulli distribution [67] | 100.0 |
| 4 | CASAS-Multiresident | NB [66] | 81.5 |
| 5 | UCI-HAR | Multiple HMMs + MOT + k-NN2 [78] | 92.6 |
| 6 | Opportunity | DLT (NMF+SAE) [28] | 99.9 |
| 7 | mHealth | HCM (2bbC + SLoSH + UOF) [81] | 97.2 |

- The highest accuracy was obtained in [81] at 97.2%, and was achieved by applying a Hierarchical Classification Method (HCM) based on the combination of binary classifiers.
- The most used classifiers in the analysed studies are SVM and CNN.

## VII. DISCUSSION

To address the discussion, Table 14 condenses the classification techniques that generated the highest accuracy for each dataset. We note that the datasets correspond to different contexts and purposes and therefore the evaluation and methods are not directly comparable between them. That is, while VanKasteren, CASAS-Kyoto, CASAS-Aruba and CASAS-Multiresident contain data captured with WSN, environmental sensors and wearables from the interactions of one or several inhabitants in indoor environments, the UCI-HAR, Opportunity and mHealth datasets contain data captured from wearables or smartphones not necessarily in indoor environments. The discussion focuses on the use of different techniques (segmentation, feature representation, feature selection, balancing or addition of instances and distribution of datasets for experimentation processes) in each dataset, in order to facilitate the decision making of researchers regarding their choice in ARS evaluation processes.

Different segmentation techniques have been applied in the VanKasteren and CASAS-Kyoto datasets in almost all the proposals. The use of time-based techniques with 60-second sliding windows and AR-SPM predominates in VanKasteren, reaching its maximum accuracy (96.7% ± 2.6) with the first mentioned technique. In the CASAS-Kyoto dataset the maximum accuracy (97.4%) was reached with the AR-SPM technique, giving a clear idea of which are the most appropriate segmentation techniques in the pre-processing of the data-streaming windows of these two datasets. In UCI-HAR [124] and mHealth [82] the segmentation for Sliding windows of samples (128 and 60 respectively) has been applied, reaching an accuracy of 91.76% and 91.9%, which have not been the maximum values of accuracy achieved. The application of segmentation techniques is still a topic to explore in these two datasets. There is only one segmentation application reference for Opportunity [16], with a sliding window of 2 seconds (1 second overlap), reaching an accuracy of 99.0%. Although it is not the maximum accuracy reached for this dataset, it is very good reference for future research. In the case of CASAS-Multiresident, there is no evidence of the application of segmentation techniques.

Regarding feature representation, most of the studies analysed for VanKasteren use the last-fired technique, obtaining the highest accuracy (96.7% ± 2.6) in [2]. In CASAS-Kyoto and mHealth there is no evidence of the application of feature representation techniques. In CASAS-Aruba, CASAS-Multiresident, UCI-HAR and Opportunity, although some techniques have been used (number of times that sensors were activated, Last-state, Change-point, z-scores and raw signals), this topic should be addressed with much greater depth since the accuracy obtained with its application has not been as significant.

Although different feature selection techniques were applied in VanKasteren (SAE, IG and PCA), the highest accuracy (96.7% ± 2.6) was obtained without the application of these techniques. For CASAS-Kyoto, a combination of CsubE and Ch2AE techniques was applied with low results and IG, with the highest accuracy at 97.4%. For CASAS-Aruba, different feature selection techniques were used (IG, PCA and CsubE + Ch2AE) with low results, the highest accuracy (100%) was obtained with statistical information about the activities. For UCI-HAR, different techniques were applied (PCA, RF, LDA and Correlation), however, the highest accuracy (92.6%) was achieved without the application of any technique. With Opportunity, the IG and SAE techniques have been used, with the latter obtaining the highest accuracy at 99.9%. For CASAS-Multiresident, the feature selection was carried out manually, reaching the maximum accuracy (81.5%) and for the mHealth dataset no feature selection technique was applied. Implementing techniques such as SAE, statistical information about the activities, IG and PCA, among others, in those datasets in which very few or no feature selection techniques have been implemented is an important challenge to be addressed.

The balancing or addition of instances in VanKasteren allowed to reach the maximum accuracy (96.7% ± 2.6) adding instances of manual form of the abnormal activities. None of the other techniques applied (SMOTE, McSI and algorithmic approach) exceeded this value. In CASAS-Kyoto only SMOTE was applied, reaching the highest accuracy of 97.4%. In Opportunity, several instance addition techniques were applied (instances randomly selected for each fold and Instance Reassignment), however the best accuracy (99.9%) was obtained without the application of techniques. There is no evidence of the application of balancing techniques or addition of instances in the tests carried out with CASAS-Aruba, CASAS-Multiresident, UCI-HAR and mHealth. It seems that the techniques of addition or balancing of instances have not been studied in depth in this field of research, preferring in some studies to add instances manually. However, there is discussion in relation to the fact that the addition of instances could generate an erroneous representation of the data really captured, as for this type of study the addition of instances by categories of activities could avoid the bias of the classifier.

The distribution of data for training and testing in the different experimentation scenarios presented in the proposals where VanKasteren and CASAS-Kyoto were evaluated was leave one day out cross-validation; in CASAS-Aruba and UCI-HAR, it was 10-fold cross- validation; in CASAS-Multiresident, it was cross-validation (3 or 10 fold); in Opportunity, it was cross-validation (5 or 10 fold) and in mHealth no such distribution was indicated.

Additionally, work must be done on the dynamic identification of window sizes according to the types of activities. Another approach that contributes considerably to this process is the use of stacked auto encoder for the automatic extraction of the features.

A relevant issue in the methodologies is that the vast majority of studies have been recreated in experimental scenarios and labelled activities, which generally yield very good accuracy rates, and are adequate as a process prior to real-time implementations. In the first hand, the approaches without real-time capabilities are used to evaluate usefully the daily activities of inhabitants in long-term within the early diagnosis of mental diseases [53]. Example of health applications for these purposes are analysing disturbed sleep cycles, which have become an indicator of mental disease such as Alzheimer, or identifying a change of patterns in activities, which is related to cognitive or physical decline.

On the second hand, the challenge of recognizing activities in real time undoubtedly generates a series of issues to address, in terms of pre-processing, in particular the representation of features and segmentation, requiring advanced methods which process data with little delay and high reliability [67]. However, the benefits of performing an activity recognition close to real time allow smart environments to provide a valuable short-term interaction with the user. For example, to promptly notify forgetfulness in the activity development, such as forgetting the umbrella a rainy day or intaking medications, or to prevent home riks of patients with dementia when manipulating household appliances, such as turning off the oven..

## VIII. CONCLUSION

The purpose of this article has been to propose a series of recommendations to researchers related to HAR, in relation to the identification of the most appropriate dataset according to the type of research. Development in this area of research in the last fifteen years has been growing from the presentation of only one research article in 2003 to 85 research articles in 2017. The most representative specialised databases, in relation to the number of scientific publications in ADL, is IEEExplore, with 66% of publications in total (247), surpassing the sum of the results obtained from consulting the databases Scopus, Science Direct, Web of Science and ACM.

It is noteworthy that the majority of publications in this field of research are proceedings or conference papers (72% of publications) and a representative percentage is in journal (27%). As for the quartile of the publication, although the highest percentage of publications is done in non-categorised proceedings or conferences (71%), a very important percentage (27%) of the publications is done in the first and second quartile journals. The USA can be highlighted as the country that accepts the largest amount of publications in this area of research, with a participation of 31.8% with respect to the total number of publications received worldwide and followed by the Netherlands with 13.1%. In terms of the number of publications generated, the USA and the UK with 13.4% of publications each, are the countries that stand out most, while the participation of Italy is also representative, with 11.2%. Although the media in which research results are published in this area of knowledge are very diverse, the most outstanding are: the magazine Pervasive and Mobile Computing (Netherlands) with 2.6% and the series of books Lecture Notes in computer science (Germany), with the same percentage. The institutions with the most experience in this field are: Ulster University (Ireland) with a participation of 4.8% of the publications worldwide and the WSU (USA) with 3.2%.

The technical analysis outlined in this paper has made it possible to identify the WSU CASAS repository as the most used, as it is referenced in 12.5% of the papers consulted, in particular with the following datasets: Tokyo, Aruba, Tulum, DOMUS, Real-Time Smart Home Stats, Single-resident apartment data and Kyoto Multiresident ADL Activities. Additionally, the VanKasteren (5.7%), UCI HAR (2.8%) and Opportunity (2.3%) datasets have shown significant representativeness in terms of usability, which allowed us to answer the hypothesis initially raised. A more detailed characterisation of the datasets referenced in the publications was made, in terms of: type of event (activities: 35.3%), occupation (single: 42.2%), annotation (annotated: 60.4%), sensing modalities (environment sensor: 46.3%).

In addition, it has been possible to identify that the vast majority of papers reviewed use a classifier based on DDA (61%). Most of the papers referred to several classification techniques for the comparative analysis of quality metrics. Accordingly, 41.7% of the references to classifiers correspond to the MM, 17.4% of the references used to the SVM classifier, while IBL is mentioned in 15.8% of the papers consulted, 15.8% of the papers mention the use of BC and 9.6% reference the use of DT.

The best results regarding the recognition of activities have been achieved using MLC, which combines several individual classifiers, as evidenced by [28], [29], [78], and [81]. In addition, the use of RNN (which should be extended to deep neural networks) considerably improves hit rates. On the other hand, proposal [67] has obtained such a high result because the Activity parameter maintains the statistical information about the activities (through Mutual Information, Frequency of triggered sensors of an activity, Interval time and Last two sensors) and by its innovative Adaptive windowing approach.

Regarding the processing of datasets with multi-occupancy, it is necessary to implement techniques that automatically select feature values, with the challenges that this entails. However, the use of genetic algorithms could be a key element to solve this challenge. Complex activity recognition (cooperative, parallel and individual) is a non-trivial problem due to the conflicts that are generated in the capture of data generated by the interactions. The rule-based approach could provide solutions to the management of such conflicts, taking into account the spatial and temporal location of the inhabitants.

## IX. FUTURE WORKS

In future work, we propose the evaluation of the recognition capacity of an ARS of complex activities, comparing the obtained quality metrics. In a first phase, we will recreate experimentation scenarios, which will be validated from our own multi-occupancy dataset and other dataset benchmarks. In a second phase, the aim is to provide real-time capabilities. In addition, we will focus on the implementation of techniques for the re-labelling of activities (not identified) and the recognition of activities using a multi-level classifier approach that integrates: 1) genetic algorithms for feature selection, and 2) Growing Hierarchical Self-Organizing Maps for classification based on the proposals of [130]–[132].

## ACKNOWLEDGMENTS

The authors would like to thank the team members (Miguel Ortiz, Sandra De-la-Hoz, Guillermo Rodriguez, Zhoe Comas, Fabio Mendoza, Dionicio Neira and Andrés Sanchez) and collaborators (Juan De-la-Hoz, Alejandro De-la-Hoz, Mario Orozco, Ernesto Esmeral, Nahum De Ávila, Walter Santiago and Jans Patiño).

## REFERENCES

[1] D. A. Umphred, *Neurological Rehabilitation*, vol. 27. Amsterdam, The Netherlands: Elsevier, no. 5, 2013.

[2] D. Arifoglu and A. Bouchachia, "Activity recognition and abnormal behaviour detection with recurrent neural networks," *Procedia Comput. Sci.*, no. 110, pp. 86–93, Jul. 2017, doi: 10.1016/j.procs.2017.06.121.

[3] M. S. Albert *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 270–279, 2011, doi: 10.1016/j.jalz.2011.03.008.

[4] F. E. Mendoza *et al.*, "Cardiovascular disease analysis using supervised and unsupervised data mining techniques," *J. Softw.*, vol. 12, no. 2, pp. 81–90, Feb. 2017, doi: 10.17706/jsw.12.2.81-90.

[5] Alzheimer's Association, "2013 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 9, no. 2, pp. 208–245, Mar. 2013, doi: 10.1016/j.jalz.2013.02.003.

[6] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "Human activity recognition from wireless sensor network data: Benchmark and software," in *Activity Recognition in Pervasive Intelligent Environments*, vol. 4. Paris, France: Atlantis Press, 2011, pp. 165–186. doi: 10.2991/978-94-91216-05-3_8.

[7] J. Ye, S. Dobson, and S. McKeever, "Situation identification techniques in pervasive computing: A review," *Pervasive Mobile Comput.*, vol. 8, no. 1, pp. 36–66, 2012, doi: 10.1016/j.pmcj.2011.01.004.

[8] A. Aztiria, J. C. Augusto, R. Basagoiti, A. Izaguirre, and D. J. Cook, "Learning frequent behaviours of the users in intelligent environments," *J. Ambient Intell. Smart Environ.*, vol. 2, no. 4, pp. 435–436, 2010, doi: 10.3233/AIS-2010-0084.

[9] M. Ghazvininejad, H. R. Rabiee, N. Pourdamghani, and P. Khanipour, "HMM based semi-supervised learning for activity recognition," in *Proc. ACM Int. Workshop Situation Activity Goal Awareness*, Sep. 2011, pp. 95–100, doi: 10.1145/2030045.2030065.

[10] P. Lasitha and S. Kodagoda, "Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features," in *Proc. IEEE 8th Conf. Ind. Electron. Appl. (ICIEA)*, Jul. 2013, pp. 567–572, doi: 10.1109/ICIEA.2013.6566433.

[11] N. Oliver, E. Horvitz, and A. Garg, "Layered representations for human activity recognition," in *Proc. 4th IEEE Int. Conf. Multimodal Interfaces*, Oct. 2002, pp. 3–8, doi: 10.1109/ICMI.2002.1166960.

[12] S. Mostafa-Al-Masum, H. Keikichi, and I. Mitsuru, "Recognition of real-world activities from environmental sound cues to create life-log," in *The Systemic Dimension of Globalization*. Rijeka, Croatia: InTech, 2011, pp. 173–190, doi: 10.5772/22491.

[13] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Recognizing independent and joint activities among multiple residents in smart environments," *J. Ambient Intell. Humanized Comput.*, vol. 1, no. 1, pp. 57–63, 2010, doi: 10.1007/s12652-009-0007-1.

[14] L. Young-Seol and C. Sung-Bae, "Activity recognition using hierarchical hidden Markov models on a smartphone with 3D accelerometer," in *Proc. Int. Conf. Hybrid Artif. Intell. Syst.*, in Lecture Notes in Computer Science, vol. 6678. Berlin, Germany: Springer, 2011, pp. 460–467, doi: 10.1007/978-3-642-21219-2_58.

[15] A. M. Mannini and A. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010, doi: 10.3390/s100201154.

[16] H. Gjoreski, M. Gams, and M. Lutrek, "Human activity recognition: From controlled lab experiments to competitive live evaluation," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 139–145, doi: 10.1109/ICDMW.2015.29.

[17] H. Gjoreski, M. Gams, and M. Luštrek, "Context-based fall detection and activity recognition using inertial and location sensors," *J. Ambient Intell. Smart Environ.*, vol. 6, no. 4, pp. 419–433, 2014, doi: 10.3233/AIS-140268.

[18] H. H. Manap, N. M. Tahir, and A. I. M. Yassin, "Anomalous gait detection based on support vector machine," in *Proc. IEEE Int. Conf. Comput. Appl. Ind. Electron.*, Dec. 2011, pp. 623–626, doi: 10.1109/ICCAIE.2011.6162209.

[19] H. Gjoreski, B. Kaluža, M. Gams, M. Radoje, and M. Luštrek, "Context-based ensemble method for human energy expenditure estimation," *Appl. Soft Comput.*, vol. 37, pp. 960–970, Dec. 2015, doi: 10.1016/j.asoc.2015.05.001.

[20] M. Altini, J. Penders, R. Vullers, and O. Amft, "Estimating energy expenditure using body-worn accelerometers: A comparison of methods, sensors number and positioning," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 219–226, Jan. 2015, doi: 10.1109/JBHI.2014.2313039.

[21] M. Gjoreski, H. Gjoreski, M. Lutrek, and M. Gams, "Automatic detection of perceived stress in campus students using smartphones," in *Proc. Int. Conf. Intell. Environ.*, Jul. 2015, pp. 132–135, doi: 10.1109/IE.2015.27.

[22] H. Alemdar, C. Tunca, and C. Ersoy, "Daily life behaviour monitoring for health assessment using machine learning: Bridging the gap between domains," *Pers. Ubiquitous Comput.*, vol. 19, no. 2, pp. 303–315, Feb. 2015, doi: 10.1007/s00779-014-0823-y.

[23] D. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, Jul. 2013, doi: 10.1109/MC.2012.328.

[24] R. Chavarriaga *et al.*, "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 2033–2042, Nov. 2013, doi: 10.1016/j.patrec.2012.12.014.

[25] N. Kawaguchi *et al.*, "HASC Challenge: Gathering large scale human activity corpus for the real-world activity understandings," in *Proc. 2nd Augmented Hum. Int. Conf. AH*, 2011, pp. 1–5, doi: 10.1145/1959826.1959853.

[26] B. Kaluža, S. Kozina, and M. Luštrek, "The activity recognition repository: Towards competitive benchmarking in ambient intelligence," in *Proc. AAAI Activity Context Represent., Techn. Lang.*, Jan. 2012, pp. 44–47.

[27] H. Gjoreski *et al.*, "Competitive live evaluations of activity-recognition systems," *IEEE Pervasive Comput.*, vol. 14, no. 1, pp. 70–77, Jan./Mar. 2015, doi: 10.1109/MPRV.2015.3.

[28] B. Chikhaoui and F. Gouineau, "Towards automatic feature extraction for activity recognition from wearable sensors: A deep learning approach," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, New Orleans, LA, USA, Nov. 2017, pp. 693–702, doi: 10.1109/ICDMW.2017.97.

[29] L. G. Fahad, S. F. Tahir, and M. Rajarajan, "Feature selection and data balancing for activity recognition in smart homes," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 512–517, doi: 10.1109/ICC.2015.7248373.

[30] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley MHAD: A comprehensive multimodal human action database," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 53–60, doi: 10.1109/WACV.2013.6474999.

[31] M. Zhang and A. A. Sawchuk, "USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, Sep. 2012, pp. 1036–1043, doi: 10.1145/2370216.2370438.

[32] B. Bruno, F. Mastrogiovanni, A. Sgorbissa, T. Vernazza, and R. Zaccaria, "Analysis of human behavior recognition algorithms based on acceleration data," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 1602–1607, doi: 10.1109/ICRA.2013.6630784.

[33] C. Chen, R. Jafari, and N. Kehtarnavaz, "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 168–172, doi: 10.1109/ICIP.2015.7350781.

[34] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognit.*, vol. 43, no. 10, pp. 3605–3620, Oct. 2010, doi: 10.1016/j.patcog.2010.04.019.

[35] M. Elhoushi, J. Georgy, A. Noureldin, and M. J. Korenberg, "A survey on approaches of motion mode recognition using sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1662–1686, Jul. 2017, doi: 10.1109/TITS.2016.2617200.

[36] X. Yang and J. Lianwen, "A naturalistic 3D acceleration-based activity dataset & benchmark evaluations," *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2010, pp. 4081–4085, doi: 10.1109/ICSMC.2010.5641790.

[37] J. M. Alcalá, J. Ureña, A. Hernández, and D. Gualda, "Assessing human activity in elderly people using non-Intrusive load monitoring," *Sensors*, vol. 17, no. 2, p. 351, Feb. 2017, doi: 10.3390/s17020351.

[38] G. Chen, A. Wang, S. Zhao, L. Liu, and C.-Y. Chang, "Latent feature learning for activity recognition using simple sensors in smart homes," *Multimedia Tools Appl.*, vol. 77, no. 12, pp. 15201–15219, Jun. 2018, doi: 10.1007/s11042-017-5100-4.

[39] T. Nef *et al.*, "Evaluation of three state-of-the-art classifiers for recognition of activities of daily living from smart home ambient data," *Sensors*, vol. 15, no. 5, pp. 11725–11740, May 2015, doi: 10.3390/s150511725.

[40] J. P. Zimmermann *et al.*, "Household electricity survey: A study of domestic electrical product usage," Intertek, London, U.K., Tech. Rep. R66141, May 2012, p. 600.

[41] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci. Data*, vol. 2, Mar. 2015, Art. no. 150007, doi: 10.1038/sdata.2015.7.

[42] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering —A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[43] C. Manterola, P. Astudillo, E. Arias, and N. Claros, "Systematic reviews of the literature: What should be known about them," *Cirugía Española*, vol. 91, no. 3, pp. 149–155, Mar. 2013, doi: 10.1016/j.ciresp.2011.07.009.

[44] L. García-Pérez *et al.*, "Systematic review of health-related utilities in Spain: the case of mental health," *Gaceta Sanitaria*, vol. 28, no. 1, pp. 77–83, May 2013, doi: 10.1016/j.gaceta.2013.04.006.

[45] C. A. Merlano-Porras and L. Gorbanev, "Health system in Colombia: A systematic review of literature," *Revista Gerencia Políticas Salud*, vol. 12, no. 24, pp. 74–86, Jan./Jun. 2013.

[46] A. Sanchez, D. Neira, and J. J. Cabello, "Frameworks applied in quality management—A systematic review," *Rev. Espacios*, vol. 37, no. 9, p. 17, Jan. 2016. [Online]. Available: http://www.revistaespacios.com/a16v37n09/16370917.html

[47] M. E. Grams *et al.*, "Validation of CKD and related conditions in existing data sets: A systematic review," *Amer. J. Kidney Diseases*, vol. 57, no. 1, pp. 44–54, Jan. 2011, doi: 10.1053/j.ajkd.2010.05.013.

[48] C. Nugent *et al.*, "An initiative for the creation of open datasets within pervasive healthcare," in *Proc. 10th EAI Int. Conf. Pervasive Comput. Technol. Healthcare*, Cancun, Mexico, May 2016, pp. 318–321, doi: 10.4108/eai.16-5-2016.2263830.

[49] S. K. Das and D. J. Cook, "Designing smart environments: A paradigm based on learning and prediction," in *Pattern Recognition and Machine Intelligence* (Lecture Notes in Computer Science), vol. 3776, S. K. Pal, S. Bandyopadhyay, and S. Biswas, Eds. Berlin, Germany: Springer, 2005, pp. 80–90, doi: 10.1007/11590316_11.

[50] ICPSR dataset. Inst. Social Res., Univ. Michigan, Ann Arbor, MI, USA. Accessed: Jul. 31, 2018. [Online]. Available: https://www.icpsr.umich.edu/icpsrweb/content/about

[51] IRBS. *International Review Boards*. Accessed: Jul. 31, 2018. [Online]. Available: https://www.icpsr.umich.edu/icpsrweb/ICPSR/irb/index.jsp

[52] N. D. Rodríguez, M. P. Cuéllar, J. Lilius, and M. D. Calvo-Flores, "A fuzzy ontology for semantic modelling and recognition of human behaviour," *Knowl.-Based Syst.*, vol. 66, pp. 46–60, Aug. 2014, doi: 10.1016/j.knosys.2014.04.016.

[53] F. J. Quesada, F. Moya, J. Medina, L. Martínez, C. Nugent, and M. Espinilla, "Generation of a partitioned dataset with single, interleave and multioccupancy daily living activities," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.* Cham, Switzerland: Springer, 2015, pp. 60–71, doi: 10.1007/978-3-319-26401-1_6.

[54] D. Cook, M. Schmitter-Edgecombe, A. Crandall, C. Sanders, and B. Thomas, "Collecting and disseminating smart home sensor data in the CASAS project," in *Proc. CHI Workshop Developing Shared Home Behav. Datasets Adv. HCI Ubiquitous Comput. Res.*, 2009, pp. 1–7.

[55] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Tracking activities in complex settings using smart environment technologies," *Int. J. Biosci. Psychiatry Technol.*, vol. 1, no. 1, pp. 25–35, Jan. 2009.

[56] *UCI Machine Learning Repository*. Accessed: Jul. 31, 2018. [Online]. Available: https://archive.ics.uci.edu/ml/index.php

[57] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "Activity recognition using semi-Markov models on real world smart home datasets," *J. Ambient Intell. Smart Environ.*, vol. 2, no. 3, pp. 311–325, Aug. 2010, doi: 10.3233/AIS-2010-0070.

[58] D. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intell. Syst.*, vol. 27, no. 1, pp. 32–38, Jan./Feb. 2012, doi: 10.1109/MIS.2010.112.

[59] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21th Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN)*, Bruges, Belgium, Apr. 2013, pp. 437–442.

[60] C. A. Ronao and S.-B. Cho, "Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models," in *Proc. 10th Int. Conf. Nat. Comput.*, Aug. 2014, pp. 681–686, doi: 10.1109/ICNC.2014.6975918.

[61] D. Roggen *et al.*, "Collecting complex activity datasets in highly rich networked sensor environments," in *Proc. 7th Int. Conf. Netw. Sens. Syst.*, Jun. 2010, pp. 233–240, doi: 10.1109/INSS.2010.5573462.

[62] P. Lukowicz *et al.*, "Recording a complex, multi modal activity data set for context recognition," in *Proc. 23th Int. Conf. Archit. Comput. Syst.*, Feb. 2010, pp. 1–6. Accessed: Jul. 31, 2018. [Online]. Available: http://www.opportunity-project.eu/challengeDataset

[63] O. Banos *et al.*, "mHealthDroid: A novel framework for agile development of mobile health applications," in *Proc. 6th Int. Workshop Conf. Ambient Assist. Living (IWAAL)*, Belfast, U.K., Dec. 2014, pp. 91–98, doi: 10.1007/978-3-319-13105-4_14.

[64] A. Shahi, J. D. Deng, and B. J. Woodford, "A streaming ensemble classifier with multi-class imbalance learning for activity recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 3983–3990.

[65] M. H. Kabir, M. R. Hoque, K. Thapa, and S.-H. Yang, "Two-layer hidden Markov model for human activity recognition in home environments," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 1, p. 4560365, 2016, doi: 10.1155/2016/4560365.

[66] H. Fang, R. Srinivasan, and D. J. Cook, "Feature selections for human activity recognition in smart home environments," *Int. J. Innov. Comput., Inf. Control*, vol. 8, no. 5B, pp. 3525–3535, May 2012.

[67] A. Shahi, B. J. Woodford, and H. Lin, "Dynamic real-time segmentation and recognition of activities using a multi-feature windowing approach," in *Proc. PAKDD Workshops*, Jeju, South Korea, vol. 10526, U. Kang, ed. Cham, Switzerland: Springer, 2017, pp. 26–38, doi: 10.1007/978-3-319-67274-8_3

[68] I. Fatima, M. Fahim, Y.-K. Lee, and S. Lee, "Effects of smart home dataset characteristics on classifiers performance for human activity recognition," in *Computer Science and Its Applications* (Lecture Notes in Electrical Engineering), vol. 203, S.-S. Yeo ed. 2012, pp. 271–281, doi: 10.1007/978-94-007-5699-1_28.

[69] N. Twomey, T. Diethe, I. Craddock, and P. Flach, "Unsupervised learning of sensor topologies for improving activity recognition in smart environments," *Neurocomputing*, vol. 234, pp. 93–106, Apr. 2017, doi: 10.1016/j.neucom.2016.12.049.

[70] N. K. Suryadevara, S. C. Mukhopadhyay, R. Wang, and R. K. Rayudu, "Forecasting the behavior of an elderly using wireless sensors data in a smart home," *Eng. Appl. Artif. Intell.*, vol. 26, no. 10, pp. 2641–2652, 2013, doi: 10.1016/j.engappai.2013.08.004.

[71] K. Amphawan, J. Soulas, and P. Lenca, "Mining top-*k* regular episodes from sensor streams," *Procedia Comput. Sci.*, no. 69, pp. 76–85, Nov. 2015, doi: 10.1016/j.procs.2015.10.008.

[72] J. W. Lee, A. Helal, Y. Sung, and K. Cho, "Context-driven control algorithms for scalable simulation of human activities in smart homes," in *Proc. IEEE 10th Int. Conf. Ubiquitous Intell. Comput.*, Dec. 2013, pp. 285–292, doi: 10.1109/UIC-ATC.2013.68.

[73] S. S. Akter and L. B. Holder, "Activity recognition using graphical features," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Dec. 2014, pp. 165–170, doi: 10.1109/ICMLA.2014.31.

[74] T. R. Bandaragoda, K. M. Ting, D. Albrecht, F. T. Liu, and J. R. Wells, "Efficient anomaly detection by isolation using nearest neighbour ensemble," in *Proc. IEEE Int. Conf. Data Mining Workshop*, Dec. 2014, pp. 698–705, doi: 10.1109/ICDMW.2014.70.

[75] T. Chanyaswad, J. M. Chang, and S. Y. Kung, "A compressive multi-kernel method for privacy-preserving machine learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4079–4086.

[76] J. Wen and Z. Wang, "Sensor-based adaptive activity recognition with dynamically available sensors," *Neurocomputing*, vol. 218, pp. 307–317, Dec. 2016, doi: 10.1016/j.neucom.2016.08.077.

[77] D. Acharjee, A. Mukherjee, J. K. Mandal, and N. Mukherjee, "Activity recognition system using inbuilt sensors of smart mobile phone and minimizing feature vectors," in *Microsystem Technologies*. Berlin, Germany: Springer-Verlag, 2015, doi: 10.1007/s00542-015-2551-2.

[78] Y.-J. Kim, Y. Kim, J. Ahn, and D. Kim, "Integrating hidden Markov models based on mixture-of-templates and k-NN$^2$ ensemble for activity recognition," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1636–1641.

[79] B. Bruno, F. Mastrogiovanni, and A. Sgorbissa, "A public domain dataset for ADL recognition using wrist-placed accelerometers," in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interact. Commun.*, Aug. 2014, pp. 738–743,

[80] O. Banos *et al.*, "Design, implementation and validation of a novel open framework for agile development of mobile health applications," *BioMed. Eng. OnLine*, vol. 14, nos. S2–S6, pp. 1–20, 2015.

[81] S. A. Khowaja, B. N. Yahya, and S.-L. Lee, "Hierarchical classification method based on selective learning of slacked hierarchy for activity recognition systems," *Expert Syst. Appl.*, vol. 88, pp. 165–177, Dec. 2017, doi: 10.1016/j.eswa.2017.06.040.

[82] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 381–388.

[83] J. Saives, C. Pianon, and G. Faraut, "Activity discovery and detection of behavioral deviations of an inhabitant from binary sensors," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1211–1224, Oct. 2015, doi: 10.1109/TASE.2015.2471842.

[84] D. I. Kim and E. Martinson, "Human centric spatial affordances for improving human activity recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 725–730, doi: 10.1109/IROS.2016.7759132.

[85] A. K. Ramakrishnan, D. Preuveneers, and Y. Berbers, "A loosely coupled and distributed Bayesian framework for multi-context recognition in dynamic ubiquitous environments," in *Proc. IEEE 10th Int. Conf. Ubiquitous Intell. Comput. and IEEE 10th Int. Conf. Autonomic Trusted Comput.*, Dec. 2013, pp. 270–277, doi: 10.1109/UIC-ATC.2013.66.

[86] R. Fallahzadeh and H. Ghasemzadeh, "Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data," in *Proc. ACM/IEEE 8th Int. Conf. Cyber-Phys. Syst. (ICCPS)*, Apr. 2017, pp. 293–302, doi: 10.1145/3055004.3055015.

[87] A. S. Billis *et al.*, "A decision-support framework for promoting independent living and ageing well," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 199–209, Jan. 2015, doi: 10.1109/JBHI.2014.2336757.

[88] E. E. Stone and M. Skubic, "Mapping kinect-based in-home gait speed to TUG time: A methodology to facilitate clinical interpretation," in *Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthcare Workshops*, May 2013, pp. 57–64, doi: 10.4108/icst.pervasivehealth.2013.252097.

[89] S. Basterrech and V. K. Ojha, "Temporal learning using echo state network for human activity recognition," in *Proc. 3rd Eur. Netw. Intell. Conf. (ENIC)*, Sep. 2016, pp. 217–223, doi: 10.1109/ENIC.2016.039.

[90] Y.-H. Chen, C.-H. Lu, K.-C. Hsu, L.-C. Fu, Y.-J. Yeh, and L.-C. Kuo, "Preference model assisted activity recognition learning in a smart home environment," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 4657–4662, doi: 10.1109/IROS.2009.5353937.

[91] J. Soulas, P. Lenca, and A. Thépaut, "Unsupervised discovery of activities of daily living characterized by their periodicity and variability," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 90–102, Oct. 2015, doi: 10.1016/j.engappai.2015.06.006.

[92] M. Ros, M. Delgado, A. Vila, H. Hagras, and A. Bilgin, "A fuzzy logic approach for learning daily human activities in an ambient intelligent environment," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jun. 2012, pp. 1–8, doi: 10.1109/FUZZ-IEEE.2012.6250770.

[93] J. Kavya and M. Geetha, "An FSM based methodology for interleaved and concurrent activity recognition," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2016, pp. 994–999, doi: 10.1109/ICACCI.2016.7732174.

[94] Y.-P. Huang and S.-R. Chen, "A fuzzy approach to discriminating heartbeat types and detecting arrhythmia," in *Proc. Int. Conf. Fuzzy Theory Appl.*, Nov. 2012, pp. 327–332, doi: 10.1109/iFUZZY.2012.6409725.

[95] N. Pathak, N. Roy, and A. Biswas, "Iterative signal separation assisted energy disaggregation," in *Proc. 6th Int. Green Sustain. Comput. Conf. (IGSC)*, Dec. 2015, pp. 1–8, doi: 10.1109/IGCC.2015.7393701.

[96] G. Acampora and A. Vitiello, "Interoperable neuro-fuzzy services for emotion-aware ambient intelligence," *Neurocomputing*, vol. 122, pp. 3–12, Dec. 2013, doi: 10.1016/j.neucom.2013.01.046.

[97] J. Shell and S. Coupland, "Improved decision making using fuzzy temporal relationships within intelligent assisted living environments," in *Proc. 7th Int. Conf. Intell. Environ.*, Jul. 2011, pp. 149–156, doi: 10.1109/IE.2011.30.

[98] N. Oukrich, A. Maach, E. Sabri, E. Mabrouk, and K. Bouchard, "Activity recognition using back-propagation algorithm and minimum redundancy feature selection method," in *Proc. 4th IEEE Int. Colloq. Inf. Sci. Technol. (CiSt)*, Oct. 2016, pp. 818–823, doi: 10.1109/CIST.2016.7805000.

[99] S. Soviany and S. Puscoci, "A hierarchical decision system for human behavioral recognition," in *Proc. 7th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2015, pp. S-79–S-84, doi: 10.1109/ECAI.2015.7301165.

[100] K. Bouchard *et al.*, "Unsupervised spatial data mining for smart homes," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 1433–1440, doi: 10.1109/ICDMW.2015.126.

[101] H. Zheng, H. Wang, and N. Black, "Human activity detection in smart home environment with self-adaptive neural networks," in *Proc. IEEE Int. Conf. Netw., Sens. Control (ICNSC)*, Apr. 2008, pp. 1505–1510, doi: 10.1109/ICNSC.2008.4525459.

[102] X. Zhang, G.-B. Kim, Y. Xia, and H.-Y. Bae, "Human activity recognition with trajectory data in multi-floor indoor environment," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, in Lecture Notes in Computer Science, Chengdu, China, vol. 7414. Berlin, Germany: Springer, 2012, pp. 257–266, doi: 10.1007/978-3-642-31900-6_33.

[103] R. C. Kumar, S. S. Bharadwaj, B. N. Sumukha, and K. George, "Human activity recognition in cognitive environments using sequential ELM," in *Proc. 2nd Int. Conf. Cognit. Comput. Inf. Process. (CCIP)*, Aug. 2016, pp. 1–6. doi: 10.1109/CCIP.2016.7802880.

[104] R. Kumar, I. Qamar, J. S. Virdi, and N. C. Krishnan, "Multi-label learning for activity recognition," in *Proc. Int. Conf. Intell. Environ.*, Jul. 2015, pp. 152–155, doi: 10.1109/IE.2015.32.

[105] C. Bhadrachalam, T. Jyothi, and T. S. Indulekha, "New approaches for discovering unsupervised human activities by mining sensor data," in *Proc. Int. Conf. Comput. Netw. Commun. (CoCoNet)*, Dec. 2015, pp. 118–123, doi: 10.1109/CoCoNet.2015.7411176.

[106] A. Aztiria, G. Farhadi, and H. Aghajan, "User behavior shift detection in ambient assisted living environments," *J. Med. Internet Res.*, vol. 1, no. 1, p. e6, Jan./Jun. 2013, doi: 10.2196/mhealth.2536.

[107] L. G. Fahad, S. F. Tahir, and M. Rajarajan, "Activity recognition in smart homes using clustering based classification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1348–1353, doi: 10.1109/ICPR.2014.241.

[108] E. Hoque and J. Stankovic, "AALO: Activity recognition in smart homes using active learning in the presence of overlapped activities," in *Proc. 6th Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth) Workshops*, May 2012, pp. 139–146, doi: 10.4108/icst.pervasivehealth.2012.248600.

[109] V. Ghasemi and A. K. Pouyan, "Activity recognition in smart homes using absolute temporal information in dynamic graphical models," in *Proc. 10th Asian Control Conf.*, May/Jun. 2015, pp. 1–6.
[110] P. Kodeswaran, R. Kokku, M. Mallick, and S. Sen, "Demultiplexing activities of daily living in IoT enabled smarthomes," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.
[111] J. L. G. Ortega, L. Han, N. Whittacker, and N. Bowring, "A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings," in *Proc. Sci. Inf. Conf.*, London, U.K., Jul. 2015, pp. 474–482.
[112] U. Avci and A. Passerini, "Improving activity recognition by segmental pattern mining," in *Proc. 8th IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2012, pp. 709–714.
[113] H. Alemdar, T. L. M. van Kasteren, M. E. Niessen, A. Merentitis, and C. Ersoy, "A unified model for human behavior modeling using a hierarchy with a variable number of states," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 3804–3809, doi: 10.1109/ICPR.2014.653.
[114] M. B. Abidine, B. Fergani, and L. Clavier, "Importance-weighted the imbalanced data for C-SVM classifier to human activity recognition," in *Proc. 8th Int. Workshop Syst., Signal Process. Appl. (WoSSPA)*, May 2013, pp. 330–335.
[115] M. B. Abidine and B. Fergani, "Evaluating a new classification method using PCA to human activity recognition," in *Proc. Int. Conf. Comput. Med. Appl. (ICCMA)*, Jan. 2013, pp. 1–4.
[116] X. Hong, C. D. Nugent, M. D. Mulvenna, S. Martin, S. Devlin, and J. G. Wallace, "Dynamic similarity-based activity detection and recognition within smart homes," *Int. J. Pervasive Comput. Commun.*, vol. 8, no. 3, pp. 264–278, 2012, doi: 10.1108/17427371211262653.
[117] V. Ghasemi, A. A. Pouyan, and M. Sharifi, "Human activity recognition in smart homes based on a difference of convex programming problem," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 1, pp. 321–344, Jan. 2017, doi: 10.3837/tiis.2017.01.017.
[118] F. A. Machot and H. C. Mayr, "Improving human activity recognition by smart windowing and spatio-temporal feature analysis," in *Proc. 9th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, 2016, p. 56, doi: 10.1145/2910674.2910697.
[119] F. A. Machot, H. C. Mayr, and S. Ranasinghe, "A windowing approach for activity recognition in sensor data streams," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2016, pp. 951–953.
[120] A. De Paola *et al.*, "An ambient intelligence system for assisted living," in *Proc. AEIT Int. Annu. Conf.*, Sep. 2017, pp. 1–6.
[121] N. Yala, B. Fergani, and A. Fleury, "Feature extractionand incremental learning to improve activity recognition on streaming data," in *Proc. IEEE Int. Conf. Evolving Adapt. Intell. Syst. (EAIS)*, Dec. 2015, pp. 1–8.
[122] R. Mohamed, T. Perumal, N. Sulaiman, N. Mustapha, and M. N. Razali, "Conflict resolution using enhanced label combination method for complex activity recognition in smart home environment," in *Proc. IEEE 6th Global Conf. Consum. Electron. (GCCE)*, Oct. 2017, pp. 1–3.
[123] S. Ntalampiras and M. Roveri, "An incremental learning mechanism for human activity recognition," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2016, pp. 1–6.
[124] C. A. Ronao and S.-B. Cho, "Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models," in *Proc. 10th Int. Conf. Natural Comput.*, Aug. 2014, pp. 681–686.
[125] E. Garcia-Ceja and R. F. Brena, "An improved three-stage classifier for activity recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 1, p. 1860003, 2018, doi: 10.1142/S0218001418600030.
[126] J. Cumin, G. Lefebvre, F. Ramparany, and J. L. Crowley, "Human activity recognition using place-based decision fusion in smart home," Arch. Ouverte HAL, Tech. Rep., 2017.
[127] G. Chetty and M. White, "Body sensor networks for human activity recognition," in *Proc. 3rd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Noida, India, Feb. 2016, pp. 660–665, doi: 10.1109/SPIN.2016.7566779.
[128] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
[129] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Proc. Int. Workshop Ambient Assist. Living (IWAAL)*, Vitoria-Gasteiz, Spain, Dec. 2012, pp. 216–223.
[130] E. de la Hoz, E. de la Hoz, A. Ortiz, J. Ortega, and A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps," *Knowl.-Based Syst.*, vol. 71, pp. 322–338, Nov. 2014, doi: 10.1016/j.knosys.2014.08.013.
[131] F. Mendoza, A. De-La-Hoz-Manotas, E. De-La-Hoz-Franco, and P. Ariza-Colpas, "Feature selection, learning metrics and dimension reduction in training and classification processes in intrusion detection systems," *J. Theor. Appl. Inf. Technol.*, vol. 82, no. 2, pp. 291–298, Dec. 2015. Accessed: Jul. 31, 2018. [Online]. Available: http://www.jatit.org/volumes/Vol82No2/12Vol82No2.pdf
[132] E. De-La-Hoz-Franco, A. Ortiz, J. Ortega, E. De-La-Hoz-Correa, and F. Mendoza, "Implementation of an intrusion detection system based on self organizing map," *J. Theor. Appl. Inf. Technol.*, vol. 71, no. 3, pp. 324–334, Jan. 2015. Accessed: Jul. 31, 2018. [Online]. Available: http://www.jatit.org/volumes/Vol71No3/2Vol71No3.pdf

**EMIRO DE-LA-HOZ-FRANCO** was born in Barranquilla, Colombia, in 1972. He received the M.Sc. degree in computer and network engineering and the Ph.D. degree in information and communication technology from the University of Granada, Spain, in 2011 in 2016, respectively. He is currently a full-time Professor and also a member of the GIECUC and the Software Engineering & Networks Research Groups, Universidad de la Costa-CUC, Barranquilla. His research interests are in the field of machine learning and recognition of activities of daily life–ADL. ORCID: 0000-0002-4926-7414.

**PAOLA ARIZA-COLPAS** was born in Barranquilla, Colombia, in 1985. She received the M.Sc. degree in computer and systems engineering from the Universidad del Norte, Barranquilla, Colombia, in 2011. She is currently pursuing the Ph.D. degree in engineering with the Universidad Pontificia Bolivariana, Medellín, Colombia. She is currently a full-time Professor and also a member of the Software Engineering & Networks Research Group, Universidad de la Costa-CUC, Barranquilla. Her research interests are in the field of data mining and image recognition.

**JAVIER MEDINA QUERO** was born in Granada, Spain, in 1983. He received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 2007 and 2010, respectively. He is currently a Researcher with the research group Intelligent Systems Based on Fuzzy Decision Analysis (Sinbad2), University of Jaén. His research interests encompass fuzzy logic, e-health, intelligent systems, ubiquitous computing, and ambient intelligence.

**MACARENA ESPINILLA** was born in Jaén, Spain, in 1983. She received the M.Sc. and Ph.D. degrees, both in computer science, from the University of Jaén, Jaén, Spain, in 2006 and 2009, respectively. She is currently an Associate Professor with the Department of Computer Systems, University of Jaén. Her current research interests include ambient intelligence, ubiquitous computing, ambient assisted living, evaluation process, decision-making, recommender system, linguistic preference modeling, and fuzzy logic-based systems.