Avances en Interacción Humano-Computadora, 2025, Volumen 10, No. 1, pp. 23-32
DOI: 10.47756/aihc.y10i1.198

RESEARCH PAPER

Datasets for Human Activity Recognition: A Comparative Analysis

Conjuntos de Datos para el Reconocimiento de Actividades Humanas: Un Análisis Comparativo

Luis F. Beltrán [1], Jessica Beltrán [2], Luis A. Castro [1,*]

Published: 30 November 2025

Abstract
Human activity recognition is a key area in human–computer interaction because it enables systems to understand people's actions and respond through implicit interaction, without requiring direct commands. This article presents a comparative analysis of six widely used public datasets, considering factors such as participant diversity, variety of activities, types of sensors, and capture conditions. The results show that although these datasets have driven the development of more accurate models, they still present limitations related to the lack of diversity and class imbalance, which affects the generalization ability of systems in real-world contexts. The need to design more inclusive and representative datasets that reflect the complexity of human interaction and support the development of fairer, more robust, and more adaptive human activity recognition systems is highlighted.

Keywords:
Human Activity Recognition; Human–Computer Interaction; Public Datasets; Data Imbalance.

Resumen
El reconocimiento de actividades humanas es un área clave en la interacción humano–computadora, ya que permite que los sistemas comprendan las acciones de las personas y respondan mediante forma implícita, sin necesidad de comandos directos. Este artículo presenta un análisis comparativo de seis conjuntos de datos públicos ampliamente utilizados, considerando factores como la diversidad de participantes, la variedad de actividades, los tipos de sensores y las condiciones de captura. Los resultados muestran que, aunque estos conjuntos de datos han impulsado el desarrollo de modelos más precisos, aún presentan limitaciones relacionadas con la falta de diversidad y el desbalance de clases, lo que afecta la capacidad de generalización de los sistemas en contextos del mundo real. Se destaca la necesidad de diseñar conjuntos de datos más inclusivos y representativos que reflejen la complejidad de la interacción humana y apoyen el desarrollo de sistemas de reconocimiento de actividades más justos, robustos y adaptativos.

Palabras clave:
Reconocimiento de Actividades Humanas; Interacción Humano–Computadora; Conjuntos de Datos Públicos; Desbalance de Datos.

1 Introduction
Human Activity Recognition (HAR) is a field of research that focuses on detecting and classifying people's actions or behaviors based on sensory data, usually captured by portable, wearable, or environmental devices. This field is closely related to Human-Computer Interaction (HCI), as the analysis of these activities enables the design of intelligent systems capable of adapting their services to the needs and contexts of users, drawing on techniques from areas such as the Internet of Things and machine learning [1]. The combination of these technologies enables the interpretation of behavior patterns and supports the development of personalized, context-aware experiences [2]. That said, data play a crucial role, as they form the basis on which the models that enable this type of recognition are trained and evaluated. Understanding the nature, quality, and structure of the datasets used is essential to ensure that systems learn in a representative and equitable manner, avoiding limitations arising from partial or biased information [3]. The analysis of the different types of datasets used in HAR represents a significant contribution to the field of HCI, as it allows us to map the current state of available resources, evaluate participant representativeness, and more accurately anticipate the performance of learning models. The validity and generalizability of these models depend on the diversity of the data. Therefore, including people of different ages, genders, contexts, and physical conditions helps to create systems that are more equitable, adaptable, and unbiased in their real-world applications [4].

Acknowledging the scope of the activities recorded in the datasets is also essential, as not all of them offer the same level of

Beltrán F. L.[1], Castro L. A.[1,*]
Instituto Tecnológico de Sonora (ITSON)
Ciudad Obregón, Sonora, México.
Email: luisf.beltran@potros.itson.edu.mx, luis.castro@acm.org

Beltrán J.[2]
Centro de Investigación en Matemáticas Aplicadas, Universidad Autónoma de Coahuila (UAdeC)
Saltillo, México.
Email: jessicabeltran@uadec.edu.mx
* Corresponding author

Avances en Interacción Humano-Computadora, 2025, Volumen 10, No. 1, pp. 23-32
DOI: 10.47756/aihc.y10i1.198

variety. Some focus on collecting simple, everyday activities [5], [6], while others capture more complex sequential activities, social interactions, or specific tasks in areas such as health, sports, or industry [7], [8]. Clearly explaining what activities are recorded, the contextual conditions under which they were captured, the types of sensors involved, and the duration of the recordings helps identify the usefulness of each dataset, recognize its limitations, and understand the challenges that still exist in applying activity recognition in underexplored environments [9].

In HAR, data imbalance arises from both natural causes and experimental design decisions. This occurs, for example, because some activities are less frequent in daily life, are short, and are underrepresented in the sample, or because temporal segmentation divides rare events [10]. It may also be due to certain participants performing a specific activity less often or to sensor failures and labeling ambiguities, which reduce the presence of rare classes. In addition to the number of instances, duration imbalance is also relevant in this field, as longer activities consume a significant part of the total time and impact the distribution of training and evaluation windows [11]. Ignoring these differences causes models to perform well with well-represented classes but fail in those that are rare and more important in practice, such as falls, health events, or anomalous behaviors [12].

Analyzing dataset imbalances enables the accurate interpretation of HAR model performance. Traditional metrics can provide a misleading view when classes are unevenly represented, so it is necessary to use balanced measures and transparently report the distribution, duration, and variability of activities [13]. Finally, addressing imbalance from a perspective of equity and human-centered design strengthens the relationship between HAR and HCI.

Based on these challenges, this work contributes by providing a systematic analysis of six widely used public datasets, examining participant diversity, activity coverage, sensor configurations, capture conditions, and different forms of class imbalance from an HCI-oriented perspective. This analysis provides a clearer understanding of the representativeness and limitations of current resources, highlighting key gaps in the data commonly used to train HAR models. The remainder of the article is structured as follows: Section 2 reviews related work; Section 3 describes the selected public datasets; Section 4 presents the comparative analysis, organized into five dimensions (participant diversity, class imbalance, activity trends, sensor types, and capture conditions); Section 5 discusses the limitations of this study; and Section 6 provides the conclusions and outlines directions for future research.

2 Related work

The recognition of human activities has been extensively explored over the last decade, driven by the growth of portable devices, embedded sensors, and human–computer interaction applications [14]. The quality and diversity of available datasets have significantly influenced progress in this field, as machine learning models rely directly on the information with which they are trained and evaluated [15]. Therefore, in recent years, there has been growing interest in analyzing the public datasets used in HAR to better understand their characteristics, limitations, and influence on model generalization [16].

Several studies have compared public datasets to identify variations in acquisition protocols, capture environments, and sensor configurations [17]. The literature highlights the introduction of the PAMAP2 dataset, which demonstrated the usefulness of integrating inertial and heart rate sensors to represent complex activities [6]. Another widely used resource is the UCI HAR Dataset, which is based on smartphone accelerometers and gyroscopes. However, it has a limited number of participants and mainly static activities [5]. The Opportunity Dataset, designed to study everyday interactions with objects [7], has also been documented, as has the HARTH dataset, which seeks to capture daily activities in real environments and include transitions between postures and movements [18].

Despite their relevance, these datasets present notable differences in their experimental design, number of subjects, session duration, and diversity of activities. Comparative studies have highlighted that the lack of homogeneity in data acquisition and capture conditions makes it difficult to compare performance among models [19]. Additionally, the number and type of sensors vary considerably. Some datasets are based exclusively on inertial sensors, while others incorporate environmental or context sensors, leading to substantial differences in the information available for learning [20].

In many datasets, the most common activities, such as walking, running, or standing, are overrepresented, while other, more critical activities, such as falls or rapid transitions, are scarce. This problem is exacerbated by temporal segmentation and individual differences among participants, generating highly skewed distributions that bias training [21]. Traditional metrics, such as overall accuracy, are often misleading in these scenarios, so it is recommended to use weighted or class-averaged measures to obtain more reliable evaluations [22]. In addition to imbalance, participant representativeness is a critical factor for the external validity of models. Several studies have shown that most available

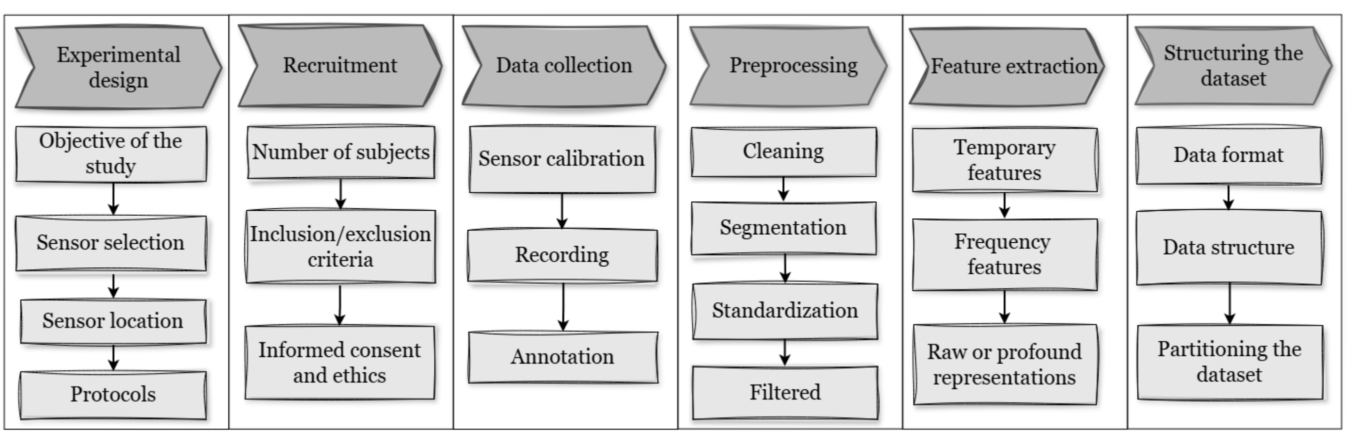

Figure 1. Phases of designing a dataset in HAR.

DOI: 10.47756/aihc.y10i1.198

datasets were obtained from small groups of subjects, generally young and healthy, which limits the applicability of the models to broader populations [15], [23]. The lack of diversity in age, gender, sociocultural contexts, and physical conditions remains a significant limitation in data collection for HAR, contributing to the reproduction of biases and reducing the general applicability of intelligent systems [24].

Various strategies have been proposed to mitigate imbalance and improve representativeness, including weighting losses by class [25], generating synthetic instances using data augmentation techniques [26], and designing partitions by subject or session to prevent information leakage [27]. However, the literature shows that these strategies are often applied in isolation, without a systematic analysis of the statistical properties of the datasets that would justify their choice.

In this context, the present work differs from previous studies by offering a comprehensive comparative analysis of the main public datasets used in HAR. Aspects such as participant diversity, activity class trends, sensor types used, and capture conditions are examined to provide a more complete picture of the current state of available resources and the challenges facing machine learning models in terms of generalization, equitable class representation, and external validity.

3 Description of public datasets

The recognition of human activities has relied heavily on public datasets that enable the evaluation and comparison of the performance of different machine learning approaches. These resources have enabled the exploration of a wide range of learning techniques, from traditional models such as SVM, k-NN, and Random Forest to deep architectures based on convolutional (CNN) or recurrent (RNN, LSTM) networks [28].

The design process for constructing a HAR dataset is illustrated in Figure 1 and consists of a series of sequential phases that ensure the dataset's quality. First, the experimental design phase establishes the study's objectives and defines the activities to be recorded, the sensors to be employed, their placement on the body, and the characteristics of the capture environment. Subsequently, the recruitment phase focuses on selecting participants who will take part in the experiment, ensuring diversity in age, gender, and physical characteristics, while also adhering to the necessary ethical considerations. Data collection involves implementing the experimental protocol, during which signals from the sensors are recorded while the subjects perform the defined activities. Once the data are obtained, the preprocessing phase is carried out, which includes noise filtering, segmentation into time windows, and normalization of the signals to reduce noise and variations that could affect the model's performance.

Subsequently, in feature extraction, relevant parameters are calculated from the signals, allowing each activity to be represented in a more discriminating manner. Finally, the structuring of the dataset organizes the information into a coherent and documented format, dividing it into training, validation, and testing subsets, facilitating its subsequent use in machine learning models and its replicability in future research.

Some of the most widely used datasets in the literature are *UCI HAR* [5], *PAMAP2* [6], *Opportunity* [7], *Sussex-Huawei Locomotion* [29], *RealWorld HAR* [8], *Skoda Mini Checkpoint* [30], Daphnet Gait [31], *MHEALTH* [32], *Activities of Daily Living* [33], *MobiAct* [34], *MotionSense* [35], among others [18], [36], [37]. In this study, we analyze six of the most commonly used datasets in the literature: UCI HAR, PAMAP2, Opportunity, MHEALTH, HARTH, and MotionSense.

The selection of these six datasets was based on specific criteria designed to ensure comparability and methodological consistency. First, we considered only datasets that are publicly available without special access restrictions and that provide standardized documentation, which facilitates reproducibility and supports consistent analysis across resources. Second, we prioritized datasets that have been widely used in the HAR literature, as evidenced by their frequent appearance in benchmark evaluations and survey studies [38]. Third, we selected datasets that offer inertial or multimodal sensor recordings with clear annotations, controlled sensor placement, and a sufficient number of participants to enable generalization analyses. Finally, we selected resources that, collectively, cover a wide variety of activity types and capture conditions, including both controlled and uncontrolled scenarios.

Below is a brief description of each one, indicating its origin, the sensors used, the type of activities recorded, and the number of participants.

3.1 UCI HAR Dataset

The UCI HAR Dataset [5] was collected at the University of California, Irvine, United States. The data comes from 30 subjects who performed six basic locomotion and posture activities: walking, ascending and descending stairs, sitting, standing, and lying down. The measurements were captured using the accelerometer and gyroscope of a Samsung Galaxy S II smartphone placed at the waist, recording three-dimensional acceleration and angular velocity signals at 50 Hz. Due to its simplicity and structured format, this dataset is widely used as a baseline reference in supervised recognition tasks.

3.2 PAMAP2 Dataset

The PAMAP2 Physical Activity Monitoring Dataset [6] was developed at the German Research Center for Artificial Intelligence (DFKI), Germany, with the aim of monitoring daily physical activity. It includes data from eight subjects equipped with three Inertial Measurement Units (IMUs) located on the wrist, chest, and ankle, as well as a heart rate sensor. It contains 12 activities that include both locomotion (walking, running, climbing stairs) and domestic and sports activities (cleaning, cycling, tennis). The signals were recorded at 100 Hz, providing high-resolution information on whole-body movement.

3.3 Opportunity Dataset

The Opportunity Activity Recognition Dataset [7] was created at ETH Zurich and TU Darmstadt for the purpose of studying complex activities in real environments. It involves four participants performing everyday tasks within an instrumented apartment, recording a total of 17 activities. The environment integrates body, environmental, and object sensors, allowing interactions with the environment to be recorded (e.g., opening doors, using utensils, or manipulating objects). It is one of the richest and most challenging datasets, featuring a comprehensive multimodal structure that combines accelerometers, gyroscopes, magnetometers, and a wide variety of state sensors.

3.4 MHEALTH Dataset

The Mobile Health (MHEALTH) Dataset [32] was developed to evaluate activity recognition in mobile health applications. Ten subjects participated in 12 physical and clinical activities, such as walking, running, jumping, strength exercises, and rehabilitation movements. A total of 12 activities were recorded, and data were collected using three inertial sensors placed on the chest, wrist, and

DOI: 10.47756/aihc.y10i1.198

Table 1. Summary of public datasets used in HAR.

Dataset	# Subjects	# Instances	Sensors	Environment	# Activities	Location	# Features	References
UCI HAR	30	10,299	A, G	Controlled	6	U.S.A.	561	Anguita et al. (2013)
PAMAP2	8	1,936,481	A, G, M	Controlled	12	Germany	54	Reiss et al. (2012)
Opportunity	4	107,803	A, G, M	Controlled	17	Switzerland	250	Roggen et al. (2010)
MHEALTH	10	343,195	A, G, M	Controlled	12	Spain	24	Banos et al. (2014)
HARTH	31	8,314,442	A	Uncontrolled	12	Norway	8	Sztyler et al. (2016)
MotionSense	24	1,412,865	A, G	Uncontrolled	6	U.K.	12	Malekzadeh et al. (2018)

A = Accelerometer, G = Gyroscope, M = Magnetometer.

ankle, along with a heart rate sensor. Signals were sampled at 50 Hz and include acceleration, angular velocity, and magnetic field intensity, offering multimodal information suitable for both fitness and clinical-monitoring applications.

3.5 HARTH Dataset

The Human Activity Recognition Trondheim (HARTH) Dataset [18] was designed to analyze activities in uncontrolled home environments. It involves 31 participants who performed 12 activities of daily living, such as walking, sitting, lying down, climbing stairs, or riding a bicycle. The data were recorded using acceleration and rotation sensors placed on the hip and leg, with a sampling frequency of 100 Hz. Its focus on naturalistic behavior makes it particularly relevant for evaluating model robustness in real-life environments.

3.6 MotionSense Dataset

The MotionSense Dataset [35] was collected at Queen Mary University of London, United Kingdom. The data were gathered from 24 subjects who performed six activities (walking, climbing and descending stairs, sitting, standing, and lying down), recorded using the accelerometer and gyroscope of an iPhone 6s placed in the front pocket of their pants. The signals were sampled at 50 Hz, with precisely annotated start and end times for each activity, offering a clean segmentation for benchmarking purposes.

4 Analysis of public datasets

This section presents a comparative analysis of the selected datasets, focusing on factors that directly affect model performance and interpretability. Rather than comparing algorithmic results, this analysis examines the diversity of participants, the nature of the activities, the types of sensors used, and the capture conditions under which the data were obtained. These aspects determine a dataset's ability to represent real-world scenarios of HCI.

To this end, we examine public HAR datasets focused on information collected through inertial sensors. Table 1 presents the six datasets selected for this analysis, detailing the number of participants, the total number of instances, the sensors used, the data collection environment (controlled or uncontrolled), the number of recorded activities, the available features, the country of origin, and the reference to the main publication.

4.1 Diversity of the participants

The diversity of participants plays a central role in ensuring that HAR models can generalize to different population groups. However, the six datasets analyzed have limited geographic and demographic coverage.

The UCI HAR dataset was collected at the University of California, Irvine, United States, with 30 young, healthy participants. PAMAP2 was recorded in Mannheim, Germany, with eight adults performing daily living and sports activities. The Opportunity dataset was developed through a collaboration between ETH Zurich (Switzerland) and TU Darmstadt (Germany), with 4 participants in an instrumented home environment. MHEALTH was captured in Spain, involving 10 healthy volunteers performing structured activities. HARTH was collected in Trondheim, Norway, with 31 subjects performing everyday activities in Uncontrolled conditions. Finally, MotionSense was recorded at Queen Mary University of London, United Kingdom, with 24 young, physically active volunteers.

Overall, four of the six datasets were collected in European institutions (Germany, Switzerland, Spain, and Norway) and one in the United Kingdom, while the remaining dataset was collected in the United States. All datasets were gathered using small, homogeneous samples composed mainly of young adults in academic environments. None include Latin American participants or reflect significant sociocultural diversity, which limits the transferability of the models to other contexts, such as Mexico or Latin America, where routines, body types, movement patterns, and infrastructure conditions may differ significantly. Therefore, it is essential to promote the creation of datasets that include participants from diverse age groups, genders, nationalities, and cultural backgrounds, thereby fostering the development of more inclusive, representative, and equitable models.

4.2 Class imbalance

The analysis of class imbalance is a central aspect in evaluating datasets used in HAR, especially when considering its impact on the interpretability and equitable performance of the resulting models. An imbalanced dataset not only affects the performance of machine learning models but can also introduce biases in the way these systems interpret and represent the diversity of human behaviors. In this article, equitable performance refers to a model's ability to behave consistently across different individuals or demographic groups, avoiding systematic advantages or disadvantages caused by the data distribution.

The following analysis seeks to describe the data structure both qualitatively and quantitatively, identifying limitations and biases before applying the datasets in machine learning and deep learning models. To this end, the distributions of instances were explored considering three levels of grouping: by subject, by activity, and by subject–activity pairs. This strategy allows us to examine not only the overall representativeness of the classes, but also the variability among participants and the differences in the execution of the same activity. Such variations are especially relevant in the context of HCI, since the perception of human activity can vary according to age, gender, physical condition, or even the capture environment, which in turn affects the system's ability to generalize to different user profiles.

Avances en Interacción Humano-Computadora, 2025, Volumen 10, No. 1, pp. 23-32
DOI: 10.47756/aihc.y10i1.198

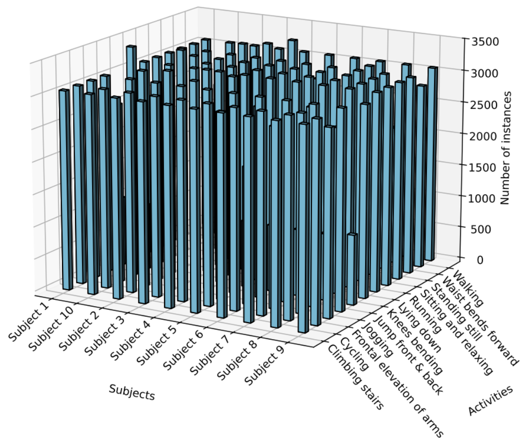

Figure 2. Class distribution per subject-activity pair on the MHEALTH dataset.

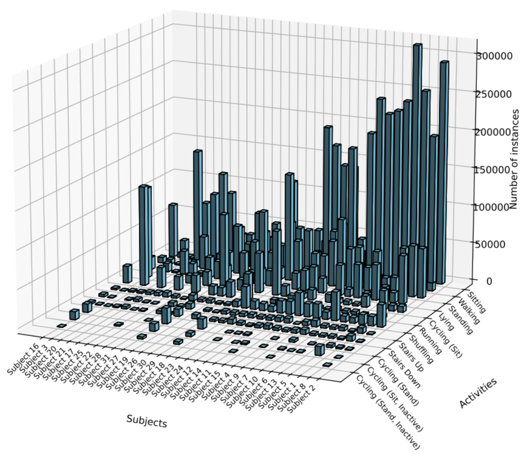

Figure 3. Class distribution per subject-activity pair on the HARTH dataset.

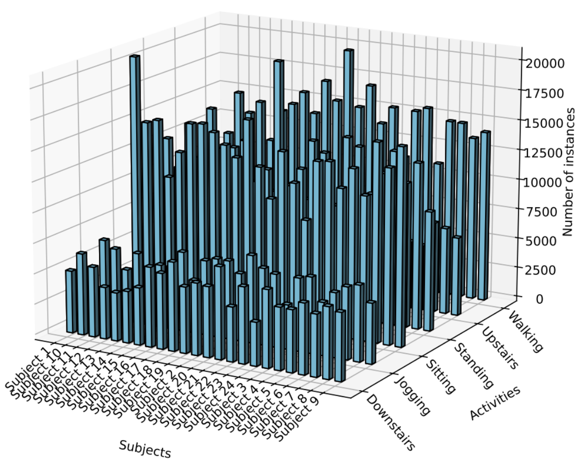

Figure 4. Class distribution per subject-activity pair on the MotionSense dataset.

The imbalance was quantified using metrics derived from information theory and descriptive statistics. First, Shannon entropy was used, defined as:

$$H = -\sum_{i=1}^{C} p_i \log_2(p_i), \tag{1}$$

where p_i represents the proportion of instances corresponding to class i and C is the total number of classes [39]. This metric captures the degree of diversity in the class distribution, reaching its maximum value when all classes are equally represented and decreasing as the imbalance increases. To allow comparisons between sets with different numbers of classes, Shannon's normalized entropy was used, calculated as:

$$H_{norm} = \frac{H}{\log_2(C)}. \tag{2}$$

In addition, other metrics were considered that provide different perspectives on distributional inequality. The Gini index measures inequality in the frequency of classes [40] and is defined as follows:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2. \tag{3}$$

Meanwhile, the dominance index quantifies the degree to which a single class predominates over the others [41], defined as:

$$D = \max_{i=1,\ldots,C}(p_i). \tag{4}$$

Finally, the coefficient of variation of the number of instances per class was calculated, a measure that relates the standard deviation σ to the mean μ of the number of instances:

$$CV = \frac{\sigma}{\mu}, \tag{5}$$

where higher values of CV reflect greater heterogeneity between classes [42].

Together, these metrics enable the comprehensive identification of imbalance patterns present in each dataset. Their analysis not only reveals differences in the distribution of activities or participants, but also provides critical information on the extent to which each dataset represents realistic scenarios of human–computer interaction, an essential aspect for the development of more robust, inclusive, and adaptive models.

By visualizing the class distributions, we can observe the heterogeneity in the identification of both dominant and underrepresented activities. Figures 2 and Figure 3 illustrate the distributions of MHEALTH and HARTH, which correspond to the lowest and highest levels of class imbalance, respectively. In contrast, Figure 4 shows the distribution of MotionSense, which is in the middle of the imbalance spectrum.

The results of the class imbalance quantification are presented in Table 2, Table 3, and Table 4, corresponding respectively to the metrics grouped by subject (MpSub), by activity (MpAct), and by subject–activity pair (MpSA). As expected, the number of MpSA combinations grows as the product of the number of subjects and activities, with HARTH reaching 372 subject–activity pairs, while Opportunity remains limited to 68. This division allows for a

Table 2. Metrics per Subject (MpSub) across datasets.

Metric	UCI HAR	HARTH	PAMAP2	Opportunity	MHEALTH	MotionSense
MpSub (N)	30	31	8	4	10	24
Entropy	4.8993	4.8696	2.989	1.9949	3.3213	4.5794
Normalized Entropy	0.9985	0.9829	0.9963	0.9975	0.9998	0.9988
GINI Index	0.9663	0.9640	0.8732	0.7482	0.8999	0.9580
Coeff. of Variation	0.1023	0.3419	0.1195	0.0842	0.0296	0.0881
Dominance Index	0.0397	0.0503	0.1407	0.2817	0.1035	0.0509

Table 3. Metrics per Activity (MpAct) across datasets.

Metric	UCI HAR	HARTH	PAMAP2	Opportunity	MHEALTH	MotionSense
MpAct (N)	6	12	12	17	12	6
Entropy	2.5759	2.5395	3.4838	3.7473	3.5498	2.4631
Normalized Entropy	0.9965	0.7084	0.9718	0.9168	0.9902	0.9529
GINI Index	0.8313	0.7658	0.9064	0.9024	0.9135	0.8060
Coeff. of Variation	0.1109	1.3454	0.3505	0.8121	0.1943	0.4053
Dominance Index	0.1888	0.3953	0.1233	0.2269	0.0895	0.2437

Table 4. Metrics per Subject-Activity pair (MpSA) across datasets.

Metric	UCI HAR	HARTH	PAMAP2	Opportunity	MHEALTH	MotionSense
MpSA (N)	180	372	96	68	120	144
Entropy	7.4658	6.8129	6.3853	5.6918	6.8686	7.0272
Normalized Entropy	0.9965	0.7978	0.9245	0.9350	0.9945	0.9801
GINI Index	0.9942	0.9860	0.9876	0.9736	0.9913	0.9917
Coeff. of Variation	0.1936	2.0523	0.7015	0.8927	0.2047	0.4430
Dominance Index	0.0092	0.0375	0.0195	0.0825	0.0100	0.0146

clearer comparison of the imbalance behavior at different granularity levels across datasets.

Using equation (1), we can see that the entropy values for MpSub are higher in UCI HAR (4.89) and HARTH (4.86), indicating greater subject diversity, while Opportunity (1.99) shows a strong imbalance. In terms of activities (MpAct), PAMAP2 (3.48) and Opportunity (3.74) stand out, suggesting a richer distribution compared to MotionSense (2.46). The normalized entropy defined in equation (2) highlights the differences more clearly: UCI HAR and MHEALTH are close to uniform distributions at all levels, while HARTH shows low normalized entropy for MpAct (0.70) and MpSA (0.79), pointing to an imbalance in activity.

Applying equation (3), we can calculate the GINI index, which corroborates the entropy results: higher values closer to 1 denote greater balance. In terms of subjects, UCI HAR and MotionSense are close to 0.96+, while Opportunity falls to 0.74, reflecting the high dominance of a few subjects. At the activity level, PAMAP2 and MHEALTH maintain a good balance (>0.90), but HARTH is weaker (0.76). This metric reveals the magnitude of the imbalance in relation to the category's size. HARTH exhibits surprising variability, particularly in MpAct (1.34) and MpSA (2.05), suggesting that certain activities and subject-activity pairs are more dominant. In contrast, UCI HAR and MHEALTH remain stable with very low variation values.

The coefficient of variation, calculated using equation (5), further highlights these patterns: HARTH and Opportunity show the greatest variability, especially in activity levels and subject-activity pairs, while UCI HAR and MHEALTH remain consistently low across all groups. PAMAP2 and MotionSense show intermediate values, with moderate variability at the activity level.

The dominance index, as shown in equation (4), measures the proportion of the largest class. HARTH again shows a marked imbalance, with dominant categories accounting for 39% of MpAct and 28% of MpSub. Opportunity also shows high dominance (28% in MpSub, 22% in MpAct). In contrast, UCI HAR maintains low dominance at all levels, with a value below 19%.

4.3 Trends in types of activities

The datasets used in HAR exhibit significant differences in the type and complexity of the activities recorded. Table 5 summarizes the main activity groups and the datasets in which they are found.

The most used datasets, such as *UCI HAR*, *MHEALTH*, and *MotionSense*, focus on locomotor and transitional activities, such as walking, climbing or descending stairs, sitting down, and standing up. These activities are common in controlled environments and allow for simple segmentation, which facilitates the comparative evaluation of algorithms. However, its scope is limited when it comes to representing the complexity of human behavior in uncontrolled contexts.

On the other hand, sets such as *PAMAP2* and *Opportunity* extend coverage to activities of daily living (ADL) and object manipulation tasks, such as cleaning, which reflect more natural and contextual interactions. Although *HARTH* includes locomotor activities, it also incorporates sports exercises and postural tasks, offering a transition between controlled and dynamic scenarios.

In general, older datasets focus on simple, repetitive activities such as walking or standing. In contrast, more recent ones incorporate

Table 5. Comparison of types of activities in the analyzed datasets.

Group	Activities	Datasets
Locomotive	Walking, running, climbing stairs, descending stairs	UCI HAR, MHEALTH, MotionSense, PAMAP2, HARTH
Transitional	Sitting, standing up, lying down	UCI HAR, MotionSense, HARTH, MHEALTH
Activities of daily living	Eating, cleaning, opening doors, using objects	Opportunity, PAMAP2
Sports or exercise	Cycling, stretching exercises, jogging	PAMAP2, MHEALTH, HARTH
Postural or rest	Lying down, standing still, sitting	UCI HAR, MHEALTH, HARTH, MotionSense

Table 6. Comparison of sensors in the analyzed datasets.

Dataset	Type	Location	Frequency
UCI HAR	Smartphone (IMU)	Waist and front pocket	50 Hz
MotionSense	Smartphone (IMU)	Front pocket	50 Hz
PAMAP2	IMUs	Wrist, chest, ankle	100 Hz
MHEALTH	IMUs	Chest, right wrist, left ankle	50 Hz
HARTH	IMUs	Chest, thigh	100 Hz
Opportunity	Multimodal	IMUs and environmental sensors	30–100 Hz

more complex and contextual activities that combine movements, everyday tasks, and object manipulation.

From 2020 to 2023, a growing trend has emerged toward capturing more realistic and context-rich human behaviors in activity recognition. Recent studies have expanded beyond isolated physical movements to include actions that reflect everyday human–technology interaction, such as walking while using a phone, performing short tasks during breaks, or multitasking in uncontrolled environments [43]. This evolution shows a shift from recognizing simple motion patterns to a broader understanding of how people behave and interact with their surroundings and personal devices, aligning with the principles of HCI.

4.4 Types of sensors used

The sensory configuration directly determines the granularity, accuracy, and ecological validity of the captured movement patterns. Table 6 summarizes the types and locations of sensors used in the analyzed datasets, highlighting how each configuration influences signal quality and contextual representativeness. Datasets such as *UCI HAR* and *MotionSense* rely primarily on smartphone-integrated sensors, typically tri-axial accelerometers and gyroscopes located in the front pocket or waistband of the user. This configuration offers portability and reflects the natural use of the device, but is more susceptible to variations due to sensor position and signal noise. Sampling frequencies in these datasets range from 50 to 100 Hz, which is sufficient to capture typical human movements, but limited for finer kinematic analysis.

In contrast, datasets such as PAMAP2, MHEALTH, and HARTH utilize multiple inertial measurement units (IMUs) distributed across different body segments, typically at the wrist, chest, and ankle, and in some cases at the thigh or lower back. This arrangement enables a more accurate reconstruction of body kinematics and a detailed representation of movement dynamics. For example, PAMAP2 utilizes three IMUs with a sampling frequency of 100 Hz, whereas MHEALTH employs three sensors synchronized at 50 Hz, recording data from accelerometers, gyros, and magnetometers. From an HCI perspective, this balance between accuracy and comfort represents a design challenge: achieving accurate measurement without compromising the usability or intrusiveness of the system. Furthermore, multimodal systems such as Opportunity combine body and environmental sensors that record not only physical movement but also contextual information about user–environment interaction. This highlights a growing trend toward hybrid sensory ecosystems that enable more adaptive and context-aware interactive systems.

4.5 Data capture conditions

Data capture conditions determine the ecological validity and generalizability of HAR datasets, as they depend on the experimental environment, sensor placement, sampling frequency, and tagging methods. In general, datasets are grouped into two types: controlled and uncontrolled. In controlled experiments, activities are performed according to a precise script under homogeneous conditions, which favors reproducibility but reduces the naturalness of movement. In the *UCI HAR*, for example, participants performed six guided activities while a Samsung Galaxy S II phone, attached to their waist, recorded accelerometer and gyroscope data at 50 Hz. The data were segmented into 2.56 s windows with 50% overlap, and labels were obtained from video recordings that validated the execution of each activity. This highly structured protocol facilitates the creation of accurate models under similar conditions but has limitations when applied in uncontrolled environments (i.e., in the wild or the real world), where device orientations and transitions between activities are more variable.

Similarly, MHEALTH used three portable sensors located on the chest, wrist, and ankle, where each activity was performed for specific time intervals defined by the experimental team, with each activity lasting between 20 seconds and 1 minute. The start and end times of each activity were marked in the experimental script, allowing the tags to be automatically generated according to the planned sequence, without the need for subsequent review by video. Before starting each session, the sensors were synchronized and checked to ensure they were correctly attached to the body, thereby avoiding displacement that could alter the measurements. The sets captured in real conditions, such as Opportunity, seek to reflect the daily behavior of users. This dataset integrates body and environmental sensors to record gestures, locomotion, and object manipulation during household tasks. Although these data include noise and spontaneous transitions that are difficult to model, they provide superior ecological validity and greater relevance for HCI applications. In the *MotionSense* dataset, participants performed six activities with an iPhone 6s in their front pocket. Although a basic script was in place, free execution introduced realistic variations without compromising experimental consistency. In summary, controlled experiments offer clean signals and accurate labeling, while uncontrolled environments provide diversity and realism at the cost of noise and complexity.

5 Limitations of the study

This study focuses exclusively on six publicly available datasets that are widely used in the HAR community. Although this selection enables a consistent and reproducible comparison, it excludes other relevant datasets whose restricted access prevents their inclusion. The analysis also depends on the information reported in the official documentation of each dataset, which in some cases lacks complete demographic or methodological details. As a result, certain aspects of diversity or contextual richness may be underrepresented. In addition, the study does not evaluate model performance on these datasets, because the objective is to examine their characteristics rather than benchmark algorithms. Finally, the HAR ecosystem continues to expand, and a broader set of datasets would allow a more comprehensive view of current trends.

6 Conclusion and future work

Analysis of the main public datasets for HAR shows that the scientific community has made significant progress in standardizing and improving the availability of data, but still faces significant challenges in terms of diversity, balance, and ecological validity. The results show that many of the most widely used datasets were designed under controlled conditions, with small and homogeneous populations, which limits their ability to generalize in uncontrolled HCI contexts. Furthermore, many existing datasets focus largely on basic locomotor and transition activities, where the technology is used solely as a sensor for non-interactive measurements, i.e., the technological device (smartphone, IMU) acts solely as a motion recorder. It's there to capture acceleration, rotation, or body position, but there is no meaningful or intentional interaction with the device's function while recording activity. This design choice severely limits the ability of data to capture and model the complex and nuanced interactions between humans and technology that are critical to the field of HCI. In addition, class imbalance persists as a structural problem that introduces bias in model training and affects its performance on less represented users or activities. From an HCI perspective, this implies that activity recognition-based systems still lack contextual sensitivity and adaptability to human diversity. Therefore, it is not enough to optimize architectures or performance metrics; a profound reconsideration of the design and collection of data that serve as the basis for learning is required. This work reinforces the idea that data quality and representativeness are as crucial as algorithmic techniques in building reliable, fair, and useful HAR systems.

This analysis opens up several lines of research aimed at strengthening the connection between HAR and HCI through more relevant data collection. First, it proposes the creation of new datasets that incorporate a greater diversity of participants, both culturally and demographically, in order to reduce the population biases observed. Second, it is necessary to develop hybrid capture strategies that combine controlled and uncontrolled environments, allowing the complexity of human activities to be recorded without losing measurement accuracy. These strategies must specifically focus on capturing dynamic human–technology interactions (e.g., multitasking, device switching, and contextual queries) rather than isolated physical movement. Finally, the results highlight the need to integrate imbalance mitigation strategies directly into machine learning pipelines, ensuring that training processes fairly represent all classes and user groups. Furthermore, future datasets should include participants from underrepresented regions, such as Latin America, and encompass more complex activities, including group or simultaneous actions, as well as explicit interactions with technology, to provide a stronger and more realistic benchmark for evaluating HAR models in modern contexts, thereby directly addressing the core research themes of HCI.

7 Acknowledgements

This work was partially funded by the Secretariat of Science, Humanities, Technology, and Innovation (SECIHTI) in Mexico, which provided a scholarship to the first author, as well as by the Sonora Institute of Technology (ITSON) through the PROFAPI program.

8 References

[1] Ann, O. C. and Theng, L. B. 2014. Human activity recognition: A review. In *2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014)*, 389–393. https://doi.org/10.1109/ICCSCE.2014.7072750

[2] Khowaja, S. A., Yahya, B. N., and Lee, S.-L. 2020. CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments. *Human-centric Computing and Information Sciences* 10, 35. https://doi.org/10.1186/s13673-020-00240-y

[3] Ye, X., Sakurai, K., Nair, N.-K. C., and Wang, K. I.-K. 2024. Machine learning techniques for sensor-based human activity recognition with data heterogeneity—A review. *Sensors* 24, 7975. https://doi.org/10.3390/s24247975

[4] De-La-Hoz-Franco, E., Ariza-Colpas, P., Quero, J. M., and Espinilla, M. 2018. Sensor-based datasets for human activity recognition—a systematic review of literature. *IEEE Access* 6, 59192–59210. https://doi.org/10.1109/ACCESS.2018.2873502

[5] Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J. L. 2013. A public domain dataset for human activity recognition using smartphones. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 437–442. https://doi.org/10.3390/s20082200

[6] Reiss, A. and Stricker, D. 2012. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th International Symposium on Wearable Computers*, 108–109. https://doi.org/10.1109/ISWC.2012.13

[7] Roggen, D. et al. 2010. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh International Conference on Networked Sensing Systems (INSS)*, 233–240. https://doi.org/10.1109/INSS.2010.5573462

[8] Sztyler, T. and Stuckenschmidt, H. 2016. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 1–9. https://doi.org/10.1109/PERCOM.2016.7456521

[9] Dentamaro, V., Gattulli, V., Impedovo, D., and Manca, F. 2024. Human activity recognition with smartphone-integrated sensors: A survey. *Expert Systems with Applications* 246, 123143. https://doi.org/10.1016/j.eswa.2024.123143

[10] Lopez-Nava, I. H., Valentín-Coronado, L. M., Garcia-Constantino, M., and Favela, J. 2020. Gait activity

classification on unbalanced data from inertial sensors using shallow and deep learning. *Sensors* 20, 4756. https://doi.org/10.3390/s20174756

[11] Cruciani, F., Cleland, I., Nugent, C., McCullagh, P., Synnes, K. and Hallberg, J. Automatic annotation for human activity recognition in free living using a smartphone. *Sensors* 18, 7 (2018), Article 2203. https://doi.org/10.3390/s18072203

[12] Zhang, J., Li, J. and Wang, W. A class-imbalanced deep learning fall detection algorithm using wearable sensors. *Sensors* 21, 19 (2021), Article 6511. https://doi.org/10.3390/s21196511

[13] Owusu-Adjei, M., Ben Hayfron-Acquah, J., Frimpong, T. and Abdul-Salaam, G. Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. *PLOS Digital Health* 2, 11 (2023), e0000290. https://doi.org/10.1371/journal.pdig.0000290

[14] Bulling, A., Blanke, U. and Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys* 46, 3 (2014), Article 33. https://doi.org/10.1145/2499621

[15] Cheng, L., Guan, Y., Zhu, K. and Li, Y. Recognition of human activities using machine learning methods with wearable sensors. In *Proc. IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC 2017)* (2017), 1–7. https://doi.org/10.1109/CCWC.2017.7868369

[16] Gu, F., Chung, M.-H., Chignell, M., Valaee, S., Zhou, B. and Liu, X. A survey on deep learning for human activity recognition. *ACM Computing Surveys* 54, 8 (2021), Article 177, 1–34. https://doi.org/10.1145/3472290

[17] Dang, L.M., Min, K., Wang, H., Piran, M.J., Lee, C.H. and Moon, H. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition* 108 (2020), Article 107561. https://doi.org/10.1016/j.patcog.2020.107561

[18] Logacjov, A., Bach, K., Kongsvold, A., Bårdstu, H.B. and Mork, P.J. HARTH: A human activity recognition dataset for machine learning. *Sensors* 21, 23 (2021), Article 7853. https://doi.org/10.3390/s21237853

[19] Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M.A. and Amft, O. A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proc. ACM Conference on Ubiquitous Computing (UbiComp 2012)* (2012), 1026–1035. https://doi.org/10.1145/2370216.2370437

[20] Benos, L., Tsaopoulos, D., Tagarakis, A.C., Kateris, D. and Bochtis, D. Optimal sensor placement and multimodal fusion for human activity recognition in agricultural tasks. *Applied Sciences* 14, 18 (2024), Article 8520. https://doi.org/10.3390/app14188520

[21] Reyes-Ortiz, J.-L., Oneto, L., Samà, A., Parra, X. and Anguita, D. Transition-aware human activity recognition using smartphones. *Neurocomputing* 171 (2016), 754–767. https://doi.org/10.1016/j.neucom.2015.07.085

[22] Lara, O.D. and Labrador, M.A. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* 15, 3 (2013), 1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192

[23] Stisen, A. et al. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *Proc. 13th ACM Conference on Embedded Networked Sensor Systems (SenSys 2015)* (2015), 127–140. https://doi.org/10.1145/2809695.2809718

[24] Jimale, A.O. and Mohd Noor, M.H. Subject variability in sensor-based activity recognition. *Journal of Ambient Intelligence and Humanized Computing* 14, 4 (2023), 3261–3274. https://doi.org/10.1007/s12652-021-03465-6

[25] Duggal, R., Freitas, S., Dhamnani, S., Chau, D.H. and Sun, J. HAR: Hardness aware reweighting for imbalanced datasets. In *Proc. IEEE International Conference on Big Data (Big Data 2021)* (2021), 735–745. https://doi.org/10.1109/BigData52589.2021.9671807

[26] Alharbi, F., Ouarbya, L. and Ward, J.A. Comparing sampling strategies for tackling imbalanced data in human activity recognition. *Sensors* 22, 4 (2022), Article 1373. https://doi.org/10.3390/s22041373

[27] Varga, D. Exposing data leakage in Wi-Fi CSI-based human action recognition: A critical analysis. *Inventions* 9, 4 (2024), Article 90. https://doi.org/10.3390/inventions9040090

[28] Dua, N., Singh, S.N., Challa, S.K., Semwal, V.B. and Sai Kumar, M.L.S. A survey on human activity recognition using deep learning techniques and wearable sensor data. In *Proc. International Conference on Machine Learning, Image Processing, Network Security and Data Sciences* (2022), 52–71. https://doi.org/10.1007/978-3-031-24352-3_5

[29] Wang, L., Gjoreski, H., Murao, K., Okita, T. and Roggen, D. Summary of the Sussex-Huawei locomotion-transportation recognition challenge. In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp 2018)* (2018), 1521–1530. https://doi.org/10.1145/3267305.3267519

[30] Plötz, T., Hammerla, N.Y. and Olivier, P. Feature learning for activity recognition in ubiquitous computing. In *Proc. 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)* (2011), 1729–1734. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-290

[31] Bächlin, M. et al. Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 436–446. https://doi.org/10.1109/TITB.2009.2036165

[32] Baños, O., Garcia, R., Holgado-Terriza, J.A., Damas, M., Pomares, H. and Rojas, I. mHealthDroid: A novel framework for agile development of mobile health applications. In *Proc. 6th International Workshop on Ambient Assisted Living and Active Ageing* (2014), 91–98. https://doi.org/10.1007/978-3-319-13105-4_14

[33] Anguita, D., Ghio, A., Oneto, L., Parra, X. and Reyes-Ortiz, J.L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Proc. International Workshop on Ambient Assisted Living* (2012), 216–223. https://doi.org/10.1007/978-3-642-35395-6_30

[34] Vavoulas, G., Chatzaki, C., Malliotakis, T., Pediaditis, M. and Tsiknakis, M. MobiAct Dataset: Recognition of Activities of Daily Living and Fall Detection Using Smartphones. In *Proc. International Conference on Information and Communication*

Technologies for Ageing Well and e-Health (2016), 143–151. https://doi.org/10.5220/0005792401430151

[35] Malekzadeh, M., Clegg, R.G., Cavallaro, A. and Haddadi, H. Mobile sensor data anonymization. In *Proc. International Conference on Internet of Things Design and Implementation (IoTDI 2018)* (2018), 49–58. https://doi.org/10.1145/3302505.3310068

[36] Sikder, N. and Nahid, A.-A. KU-HAR: An open dataset for heterogeneous human activity recognition. *Pattern Recognition Letters* 146 (2021), 46–54. https://doi.org/10.1016/j.patrec.2021.02.024

[37] Weiss, G.M. WISDM smartphone and smartwatch activity and biometrics dataset. *UCI Machine Learning Repository* (2019). https://doi.org/10.24432/C5HK59

[38] Alam, G., McChesney, I., Nicholl, P. and Rafferty, J. Open datasets in human activity recognition research—Issues and challenges: A review. *IEEE Sensors Journal* 23, 22 (2023), 26952–26980. https://doi.org/10.1109/JSEN.2023.3317645

[39] Ali, A., Anam, S. and Ahmed, M.M. Shannon entropy in artificial intelligence and its applications based on information theory. *Journal of Applied Emerging Sciences* 13, 1 (2023), 9–17.

[40] Hayat, A., Morgado-Dias, F., Bhuyan, B.P. and Tomar, R. Human activity recognition for elderly people using machine and deep learning approaches. *Information* 13, 6 (2022), Article 275. https://doi.org/10.3390/info13060275

[41] Magurran, A.E. *Measuring biological diversity.* John Wiley & Sons (2013).

[42] Abdi, H. Coefficient of variation. In *Encyclopedia of Research Design*, Vol. 1 (2010), 169–171. https://doi.org/10.4135/9781412961288.n56

[43] Bouton-Bessac, E., Meegahapola, L. and Gatica-Perez, D. Your Day in Your Pocket: Complex Activity Recognition from Smartphone Accelerometers. In *Proc. International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth 2022)* (2022), 247–258. https://doi.org/10.1007/978-3-031-34586-9_17