RH-IoT-1: A Real-World Smart-Home Dataset for Room-Level Human Presence Detection

Sehrish Rafique[1,2], Patrick Holthaus[1], Gu Fang[2] and Farshid Amirabdollahian[1]

[1] Robotics Research Group, University of Hertfordshire, Hatfield, United Kingdom
{s.rafique, p.holthaus, f.amirabdollahian2}@herts.ac.uk
[2] School of Engineering, Design and Built Environment, Western Sydney University, Sydney, Australia
22200820@student.westernsydney.edu.au, g.fang@westernsydney.edu.au

Abstract. As home-based care continues to expand after hospital discharge, there is an increasing need for indoor tracking, so assistive robots know where the person to be served is. However, current smart home datasets fail to reflect realistic home environments for training multi-room tracking and transitions. To address this gap, we introduce the Robot House IoT Dataset (RH-IoT-1), a real-world, multi-sensor smart home dataset to support research on room-level human presence detection for healthcare and robotics. The dataset consists of 68 heterogeneous sensors, including Passive Infrared (PIR) motion sensors, door contacts, appliance monitors, light sensors, and seat occupancy sensors, from a fully instrumented domestic environment. To demonstrate the utility of this dataset, we further present a baseline evaluation for room-level presence detection using a Random Forest (RF) classifier, achieving up to 99.15% accuracy under controlled conditions and above 91% on unseen data. These results demonstrate that accurate room-level presence detection is achievable relying solely on ambient sensor data. RH-IoT-1 provides a realistic benchmark for developing privacy-aware localisation methods in smart homes and supports future research in assistive robotics, Ambient Assisted Living (AAL), and hospital-at-home environments.

Keywords: Social Robotics, Human Presence Detection, Ambient Assisted Living, Multi-Sensor Dataset, Human–Robot Interaction

1 Introduction

With a global shift towards an increasingly older population [1], the demand for healthcare models that support safe and independent living at home is also accelerating. The growth of diagnoses in chronic diseases, specifically in older populations [2], has further intensified this need [3]. In such settings, people can benefit from the advantages of their own home [4], while being professionally cared for and supported by ambient technologies or assistive robots [5,6]. Recent research consistently shows that for such systems to operate safely, they require a key component: continuous awareness of the resident's location and context within the house [7,8,9,10]. Continuous awareness of location enables detection of prolonged inactivity, risk monitoring, safe and meaningful robot assistance, and coordination with care providers [11].

As a part of this approach, assistive robots have received increasing attention as they are physically capable of delivering medication, helping with activities of daily living, and proactively approaching users, for example, to remind them of appointments [12]. For these robots to interact efficiently, they need a strong awareness of their surroundings and where the person is [13]. Recent research in social robotics suggests that assistive robots should determine when assistance is appropriate rather than jumping to help inappropriately at the wrong locations and times. To support such decisions, robots require reliable contextual information about where the person is within the home. Room-level localisation provides this context; e.g. prolonged bathroom presence may indicate a potential problem, whereas extended time in a living room may reflect normal behaviour.

Assistive robots can localise people in their home using WiFi, Bluetooth, and vision-based technologies [14]; however, these approaches often raise privacy concerns or require significant computational power to process. Ambient sensor-based presence detection and localisation offer a privacy-aware and low-cost alternative [15]. Ambient sensing enables continuous, unobtrusive monitoring in the background, without asking people to wear devices or deal with intrusive data collection. This makes it especially well-suited for private spaces like bedrooms and bathrooms, where comfort and dignity really matter. Yet despite growing interest in smart-home healthcare systems [16], only a limited number of datasets support privacy-preserving and affordable room-level presence detection. Existing datasets in this domain often focus on activity recognition only, rather than including spatial localisation and fully labelled room transitions [3,17,18,19,17,20]. Critically, the majority of these datasets are not collected in realistic home environments. Instead, they are mostly scripted and collected on artificial testbeds that do not sufficiently resemble real homes. In addition, only a small number of existing datasets [21] are designed to support localisation tasks required by assistive robots, as most of them mainly focus on recognising activities and do not include clear location information or continuous tracking of rooms. To address these limitations, we collected a new dataset, Robot House IoT Dataset (RH-IoT-1), in the University of Hertfordshire's Robot House (RH), a typical British home in a residential area that has been adopted for Human-Robot Interaction (HRI) research. This dataset is specifically designed to support room-level human presence and context awareness for robot assistance and thereby helps to connect basic sensing with higher-level reasoning. It allows robots to base decisions about offering help on real environmental and behavioural evidence.

This paper provides two key contributions for the social robotics community, aimed at supporting the development of assistive robotics and Ambient Assisted Living (AAL). First, we present RH-IoT-1, a novel, real-world multivariate smart home dataset, consisting of 68 sensor inputs containing 63,480 time-stamped samples. It consists of data covering a variety of commonly available ambient sensors, including Passive Infrared (PIR) motion sensors, contact sensors, pressure mats, temperature monitoring sensors, brightness sensors and appliance sensors, collected over five days with a single person. Each observation is labelled manually using the video-recorded data to provide a reliable ground truth. The provided annotations ensure that room transitions are accurately reflected to support precise room-level presence detection, i.e. that when a person moves from one room to another, this transition is explicitly represented. These labels capture changes that often occur within a few seconds. This distinction is important because the models trained on accurately labelled room data help assistive robots to understand the surroundings first, before deciding on the assistance. Second, we provide an initial baseline analysis demonstrating the feasibility of presence detection using ambient sensors. We used a Random Forest (RF) model as a baseline, and it achieved over 99.15% accuracy on the held-out test set, using an 80/20 train–test split. When evaluated on unseen data (not used during training and testing), performance remains above 91%. These results demonstrate the usefulness of RH-IoT-1 in AAL applications and attest to its strong potential to be used for decision-making in assistive robots.

The remainder of this paper is organised as follows. Section 2 reviews existing datasets and identifies limitations for application in assisted living scenarios with interactive robots. Section 3 introduces our proposed dataset RH-IoT-1, describing its characteristics and collection process. Section 4 presents a comparison of relevant benchmark datasets for comparison purposes. Section 5 shows a baseline model and experimental setup used to confirm the dataset's general usefulness. Section 6 discusses potential applications of our dataset in assisted living scenarios, and Section 7 shows the limitations of the proposed dataset. Finally, Section 8 concludes the paper and outlines directions for future work.

2 Related Work

Human presence detection and localisation within domestic environments play a central role in AAL and assistive robotics. AAL and assistive robots often rely on knowing where a person is inside the home in order to provide meaningful support. For example, [22] presents a system in which a mobile robot monitors residents, navigates through the home, and enables communication with caregivers. For these tasks, the robot needs to determine the user's location so it can move to the correct place and provide assistance when needed. Similarly, the MoveCare project described in [23] explores long-term deployment of an assistive robot in domestic environments to support monitoring, interaction, and daily assistance for elderly users. In such settings, localisation provides important context, allowing the system to identify where help may be required or where the user can be found. More broadly, smart home environments designed for assistive technologies, such as those discussed in [24], combine sensing and robotic systems to support residents across different areas of the home.

Several techniques have been studied for indoor human presence detection and localisation. These techniques include WiFi signals, Bluetooth, smartphones, and radiofrequency signals [25,26,27]. However, the majority of these require additional hardware apart from pre-installed sensors in smart homes and may introduce privacy concerns and deployment complexity. In contrast, purely ambient sensors based room level localisation remains comparatively less explored, while being more affordable and more privacy-preserving. The majority of ambient sensing-based existing smart home datasets, such as CASAS [3], ARAS [18], TUM Kitchen [19], Kasteren [17], KU-HAR [20], and UniMiB SHAR [28], are not well-suited for developing localisation-based assistive applications, as they primarily focus on activity recognition only. Only limited work has explored ambient sensing datasets that combine room-level presence detection with explicit spatial ground truth to support robotic assistance. The lack of explicit room-level ground truth limits progress in context-aware assistive robotics. Without spatial information, systems struggle to move beyond basic activity recognition. It becomes harder to understand where events are taking place inside the home. Consequently, there remains a need for datasets that enable privacy-preserving localisation in real domestic environments. The RH-IoT-1 dataset introduced in Section 3 is developed to address this gap directly. By providing reliable room-level annotations on ambient sensor data, it supports research that integrates spatial and behavioural understanding.

3 RH-IoT-1 Dataset

As outlined in Section 2, no available dataset provides detailed information for room-level human presence detection with precise transition labels. Here, we describe the RH-IoT-1 dataset[3] with fully labelled transitions under real-world conditions, starting with characterising the environment and sensor placement strategy used for obtaining the dataset. We will then provide further information on data acquisition, annotation, and other preprocessing steps applied to the raw data to allow for the collection of compatible datasets by other researchers.

3.1 Dataset Environment

This dataset consists of sensory data from the University of Hertfordshire's Robot House (RH)[4], a four-bedroom British residential home. The ground floor of the RH is adapted to conduct research and studies in AAL scenarios for HRI and social assistive robotics solutions. As an off-campus facility, the environment reflects realistic home settings, where participants could move freely, as if they were living in their own home, with opportunities for sleeping, cooking, doing dishes, laundry, watching TV, etc., in appropriate places like the bedroom, kitchen, bathroom, living area, and dining area. Figure 1 shows the layout of

[3] Request access: `https://robothouse-dev.herts.ac.uk/datasets/RH-IoT/1/`

[4] Further information about the facility: `https://robothouse.herts.ac.uk/`

RH's ground floor. The entire ground floor was available during the experiment. The conservatory and office are excluded from the dataset as they serve as observation decks for conducting experiments. Since the living room functions as both a 'TV and sofa lounge' and a 'dining space', we divided it into two separate zones (Dining Area and Sofa Area). Between the sofa area and the corridor, there is a small staircase. As a result, the seven areas selected for our experiment were the dining area, the seating area within the living room (called the sofa area), the kitchen, the bathroom, the corridor, the staircase, and the bedroom.

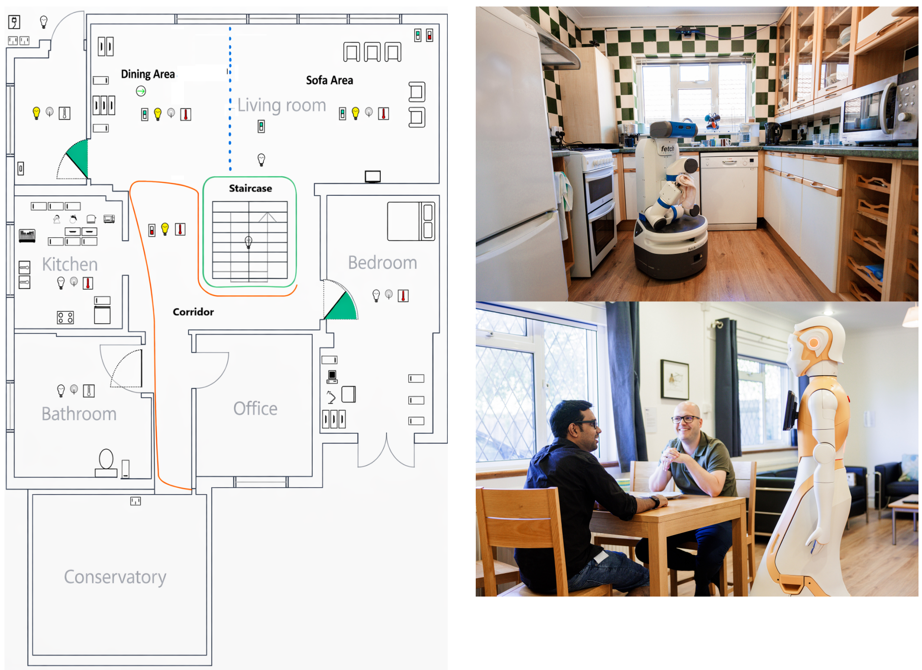

Fig. 1. (a) Illustrative floor plan of the Robot House, indicating where the dataset sensors are placed across rooms. (b) Assistive scenario in the kitchen, where a robot fetches a towel for a user. (c) Example human–robot interaction where a robot provides support to two people in the Robot House.

3.2 Sensor Placement

Each room in the RH is equipped with ambient sensors. These sensors include PIR motion sensors, contact sensors on doors, drawers, toilet lid, and fridge door, plug monitors on microwave, kettle, and coffee machine, pressure mats on sofa and bed. Additional environmental sensors are used for the brightness and water flow usage monitoring.

Table 1. Power consumption examples of monitored appliances

Device State	STATUS	VALUE
Microwave plugged in (idle)	1	1.9
Microwave running	1	1523.5
TV plugged in (screen off)	1	1
TV active	1	187.8–197.9

We used all of these sensors in our analysis, rather than using only a single type of sensor. This was important, since, for example, a PIR motion sensor alone is insufficient for accurate room-level presence detection. They often miss still activities like sitting or sleeping, and also struggle to distinguish between movement in nearby rooms, which creates confusion. By integrating multiple sensor types, we could capture both active and stationary behaviours, offering a more comprehensive representation of occupancy. The inclusion of object-usage information further improves contextual understanding and reduces confusion between neighbouring spaces.

3.3 Sensor Data Characteristics

The dataset consists of multivariate time-series data collected from RH, equipped with a diverse set of ambient sensors. In total, 68 unique sensors were recorded. In the processed version, each sensor provides four data channels: **STATUS, VALUE, XCOORD, and YCOORD**. This results in 272(68 sensors$\times$4 attributes) sensor feature streams in total, excluding timestamp, location, and activity labels. The **XCOORD** and **YCOORD** are X and Y coordinates of the sensor's location within the house. The **STATUS** field shows the basic binary state of sensors in binary values such as **0** (inactive, Closed, Off, Occupied, FALSE, Empty, Absent, Inconvenient) or **1** (active, Open, On, Free, TRUE, Good, Present, Comfortable). The **VALUE** field shows the actual measurement. For binary sensors, the value is usually the same as the status. For continuous sensors like voltages, brightness or water flow, the value gives the real measurement. So, numeric sensors with status and value together can look like the examples given in Table 1. Of the 68 unique sensors, 33 are binary. The remaining 35 sensors produce continuous numerical measurements. These include brightness levels, water flow readings, etc. Further details on sensor types and characteristics are provided in the accompanying README file.

3.4 Data Acquisition

Sensor data were collected using a modular system developed for the RH setup. Each sensor was polled in a separate thread at approximately 0.2-second intervals. However, timestamps were recorded at one-second resolution, meaning multiple readings may share the same timestamp, resulting in quasi-regular sampling rather than perfectly uniform. The median inter-sample interval is 1.0 sec-

Table 2. Sensor Categories and Coverage

Category	Sensors (Count)	Covered Sensors
PIR Motion Sensors	6	Bathroom, Bedroom, Corridor, Dining, Kitchen, Sofa
Door/Contact	25	Bathroom Door, Bedroom Door, Fridge Door, Ceiling Cupboards (3), Floor Cupboards (3), Small Cupboards (2), Floor Drawers (2), Small Drawers (3)
Seat/Pressure	7	Bed, Sofa Seat 0–4, Table Pressure 2
Appliances/Power	14	Kettle, Coffee Machine, TV, Toaster, Fridge, Microwave (if present), plugs (various)
Lamp	5	Bedside Lamp, Dining Lamp, Sofa Lamp, Central Lamp, Table Lamp
Brightness	6	Bathroom, Bedroom, Corridor, Dining, Kitchen, Sofa
Water Flow	2	Sink Cold, Sink Hot
Toilet Flush	1	Toilet Flush
TV Activity	2	TV, watchingtv
Total	68	All sensors in dataset

onds. Data were saved hourly as JSON files. For this release, the hourly files were merged, chronologically ordered, and exported as a single CSV file.

3.5 Data Annotation Protocol

Room annotations were manually created by a single annotator by carefully reviewing synchronised video recordings. The videos were used only to confirm the ground truth and were not included in the model's training process. We decided to annotate everything manually to ensure high accuracy, particularly for identifying the transitions. Extra attention was given when identifying transitions between rooms to make sure the timing was consistent, and the boundaries were marked as accurately as possible. Automated labelling approaches were considered, but they were not suitable for this study due to their tendency to introduce errors and their limited reliability.

The annotated dataset contains 7 distinct room-level locations. The room labels used in this dataset are **kitchen, bedroom, bathroom, corridor, staircase, sofa area and dining area**. The dataset also includes 13 activity labels: **cooking, dishes, having a meal, having tea, meal preparation, prayers, preparing a hot drink, sitting, standing, toileting, walking, watching TV**, and an **other** category. To address severe class imbalance in the baseline models, any activity with fewer than 300 instances was combined into the single **other** group. The other category includes hygiene (121 instances), resting (89 instances), and preparing cold drinks (85 instances).

3.6 Data Preprocessing

To keep all sensor streams aligned in time, we resampled the data to a common 1 Hz timeline. As expected, this introduced gaps of approximately 12.49% in the overall data. We filled the missing sensor values using forward filling. For numerical sensors representing intensity-based measurements (e.g., brightness or water flow), the data were then scaled using min–max normalisation. This helped bring all sensors onto a similar range, making comparisons more meaningful.

4 Comparison of RH-IoT with Benchmark Datasets

Several benchmark datasets of ambient sensing have been selected for comparison, including CASAS [3], ARAS [18], TUM Kitchen [19], Kasteren [17], KU-HAR [20], and UniMiB SHAR [28]. Table 3 provides a comparison of these datasets to find the relevance of the proposed presence detection and localisation tasks. Only three of these datasets are analysed in detail based on their relevance to ambient sensing and real-world annotation availability. The following subsections evaluate three closely relevant datasets to determine their usability for presence detection and localisation tasks, and how the RH-IoT-1 dataset has been more helpful.

Table 3. Comparison of Ambient Sensing-Based Benchmark Datasets to Determine Usability for Human Presence Detection and Localisation

Dataset	Sensors	Sensor ID	Locatio	Rooms	Values	Labels	Annotations
CASAS [3]	PIR, door, temp., power	Yes	Yes	Yes	Binary	Daily activities	Time-stamped, multi-resident
ARAS [18]	PIR, door, temp.	Yes	Yes	Yes	Binary / numeric	Activities	Multi-resident, time-stamped
KU-HAR [20]	Accelerometer, gyro (wearable)	Yes	NA	NA	Continuous	Walking, sitting, standing	Smartphone sensor dataset
Kasteren [17]	PIR, door, pressure	Yes	Yes	Yes	Binary	Daily activities	Time-stamped single resident
UniMiB SHAR [28]	Accelerometer, gyro (wearable)	Yes	NA	NA	Continuous	Walking, jogging, sitting	Smartphone activity dataset
RH-IoT v1	PIR, door, light, appliance, pressure, environment (68)	Yes	Yes	7	Binary + numeric	Room presence + 13 activities	Video-verified ground truth

ARAS Dataset The ARAS dataset [18] is a low privacy-revealing dataset, containing only binary sensors, collected in two homes with participants of different ages. It has a total of 27 attributes, and the dataset was collected at 10 hertz. The activities were performed in a completely natural way. The ground truth was collected using GUI based self reporting mechanism, so participants could easily input the activity they were performing. At the time of final labelling, the previous activity was assumed and repeated until the next annotation was provided. If this dataset is used for presence detection, the assumption-based labelling may affect precision, which is sensitive to timing and activity transitions. This makes this dataset suitable for only activity recognition with low privacy exposure. However, for localisation, finer spatial and transition-aware information is needed.

CASAS Dataset CASAS dataset [3] provides a variety of datasets collected in real-world scenarios, properly annotated for machine and deep learning tasks focused on human behaviours. These datasets contain timestamps, sensor IDs, sensor status, sensor values, and activity labels. These ambient sensor datasets are specifically for Human Activity Recognition (HAR) and lack precision for locating or finding people. Moreover, localisation requires temporal data for detailed contextual analysis, which is missing in the CASAS datasets.

Kasteren Dataset The Kasteren dataset [17] is collected from a single resident in a home over several weeks. A simple sensor setup is used for data collection, including PIR motion sensors and contact sensors on appliances and doors. It provides sensor IDs, room locations, and activity transitions for eight activities. However, it also lacks detailed spatial and temporal annotations required for presence detection tasks.

Proposed RH-IoT-1 In comparison to the CASAS [3], ARAS [18], and Kasteren [17] datasets, the RH-IoT-1 provides finer-grained annotations for room-level presence detection, particularly due to the inclusion of pressure sensors and explicitly labelled transitions. It captures spatial information through sensor coordinates, allowing more precise identification of occupancy and movement between areas. The dataset also includes detailed ground truth on room transitions, which further supports the development of presence detection and localisation approaches for assistive robots within this environment.

5 Baseline Model

To assess the utility value of RH-IoT-1, i.e. determine if it is suitable for simple classification tasks, we conducted a baseline experiment for room-level presence detection using an RF classifier on a subset of the RH-IoT-1. The baseline subset includes 39 ambient sensors: motion sensors (6), door and cupboard contact sensors (17), appliance sensors (7), pressure sensors on the bed and sofa seats (6), and utility sensors for toilet flush and water pipes (3). Some sensors that were not actively used were removed. Although these sensors belong to the defined categories in the dataset, they were not required for this specific model. Other researchers can still use the complete dataset depending on the needs of their chosen models.

The reason for removing these sensors was based on the analysis of the participant's personalised routine. For example, among the five sofa seats in the home, a person living alone may consistently sit on seat 2, while the remaining seats are rarely or never used. Similarly, a person who does not drink coffee may never use the coffee machine. The spatial coordinates (x and y) available in the RH-IoT-1 dataset were also not used in the baseline experiment. This allows the model to rely solely on ambient sensor signals, while the spatial information is also valuable and will be used in more detailed experiments when integrating presence detection with robot navigation tasks.

RF was chosen because it handles imbalanced data well, performs strongly on structured sensor inputs, and allows us to interpret results through feature importance. The model used 100 decision trees, each with a maximum depth of 10. A node could only be split if it contained at least 5 samples, and each leaf node required a minimum of 2 samples. During each split, only a random subset of features was considered (a technique known as feature bagging).

Room labels were converted into numerical form using `LabelEncoder`. Each tree produced its own prediction, and the final classification was determined by majority voting across all trees.

$$\hat{y} = \operatorname{mode}\left(h_1(x), h_2(x), \ldots, h_T(x)\right) \tag{1}$$

where $\hat{y}$ denotes the predicted room label, $h_t(x)$ is the prediction of the t-th decision tree, and $T = 100$ is the total number of trees.

5.1 Experimental Scenarios

To better understand how different sensing configurations affect room-level presence detection, we evaluated the model under three progressively constrained scenarios. These scenarios simulate varying levels of sensing availability. This allows us to assess how robust the model remains when sensing information becomes limited. The progressively constrained scenarios to reflect different sensing conditions are:

- **Scenario 1:** Full subset of sensors, all locations.
- **Scenario 2:** Full subset of sensors, but sparse locations(corridor and staircase) and their sensors excluded.
- **Scenario 3:** PIR motion sensors only, excluding staircase (since it has no PIR motion sensor)

Performance was evaluated using accuracy and weighted F1-score.

5.2 Train/Test Strategy

The dataset was split using an 80/20 train–test split. In addition, we evaluated the model on a separate unseen test dataset (that was not used during training or testing) in order to assess generalisation.

5.3 Results

Under the 80/20 train–test split, RF achieved:

- **Scenario 1:** 95.00% accuracy (Weighted F1 = 0.94)
- **Scenario 2:** 99.15% accuracy (Weighted F1 = 0.99)
- **Scenario 3:** 78.98% accuracy (Weighted F1 = 0.79) ,

Results on unseen test data: The results remained stable:

- **Scenario 1:** 91.43% accuracy (Weighted F1 = 0.9189)
- **Scenario 2:** 90.53% accuracy (Weighted F1 = 0.8939)
- **Scenario 3:** 82.04% accuracy (Weighted F1 = 0.7955)

The results suggest that room-level human presence can be reliably inferred from standard ambient sensors without the need for extra hardware or intrusive sensing methods. While the model achieved high accuracy under the random train–test split, the evaluation on unseen test data provides a more realistic estimate of performance. Results remain reasonably consistent across unseen data, and the model maintained strong class separation and balanced predictions across most rooms. Sparse classes such as *staircase* remained challenging due to limited training instances. Misclassifications mainly occurred between neighbouring areas, such as the corridor and dining area, which is expected, given the natural overlap in movement patterns near transitional boundaries.

5.4 Feature Importance and Interpretability

We analysed the importance of the chosen sensors using RF feature importance analysis. The feature importance analysis suggests that PIR motion sensors were some of the strongest indicators of room-level presence, particularly in transitional spaces such as the corridor and kitchen. Pressure mats were especially helpful for detecting stationary activities like sitting and resting. In addition, contact-based sensors, including toilet flush and door sensors, played a key role in identifying bathroom and bedroom usage. These findings highlight how different ambient sensors complement one another. Knowing which sensors drive predictions is particularly valuable in assistive robotics and home-based care environments such as Hospital At Home (H@H), where trust and transparency are as important as accuracy.

6 Applications in Assisted Living

The RH-IoT-1 dataset is designed to support research in AAL. In particular, it is designed to help with spatial awareness and assistive robotics inside real home environments. Rather than focusing only on raw sensor processing, the dataset supports research on intelligent home support systems. We specifically foresee the following applications as potential use cases.

6.1 Room-Level Presence Detection for Assistive Robots

Room-level presence detection acts as an intermediate layer between raw sensor data and robot navigation within the home. Instead of searching the entire home, a service robot can narrow its scope to a single room where presence is detected using models developed with the RH-IoT-1 dataset. This reduces search time and unnecessary movements. Room-level presence detection acts as a gating mechanism for interaction with people to be served. Before initiating a dialogue, delivering medications or offering any assistance, a robot must confirm the correct spatial zone. Social and service rules differ across rooms. The space or environment around the robot also limits and shapes the behaviour and decision-making. What is appropriate to be served in the living room may be inappropriate and intrusive in the bathroom or bedroom. By providing clear room-level information, the RH-IoT-1 dataset helps researchers design assistive robots that adjust their behaviour according to the space they are in.

6.2 Context-Aware Activity Recognition

This dataset lets researchers look at both activity and location together. They can see what people do in different rooms and how they move from one space to another. For example, knowing that a person is 'sitting' is good and provides a useful piece of information. But it will be more informative if it can tell that the person is sitting on the sofa while watching TV for 3 hours. It becomes more useful when combined with the spatial context. The RH-IoT-1 dataset provides the opportunity to conduct such analyses.

6.3 Infrastructure for Privacy-Preserving Intelligent Homes

RH-IoT-1 relies entirely on ambient sensors, avoiding any intrusive methods like cameras and microphones. This decision reflects practical acceptance constraints in domestic healthcare settings, particularly in private spaces such as bathrooms and bedrooms, where residents require full privacy. By avoiding vision and audio recording, the dataset supports the development of assistive robots that balance services with user privacy. It also sets the baseline for studying how to build efficient models that can run on robots with limited computing power. Moreover, the RH-IoT-1 dataset serves as a key resource for advancing assistive robotics in H@H [29] settings. H@H models aim to deliver hospital-level care within patients' homes, particularly for older adults, reducing hospital admissions and lengths of stay while improving patient satisfaction. In such settings, privacy-preserving sensing and intelligent assistive systems can support continuous monitoring and early intervention without compromising residents' comfort.

7 Dataset Limitations

Although the RH-IoT-1 dataset provides high-resolution room-level annotations and a diverse set of ambient sensors, it is important to acknowledge several limitations. Firstly, the data was collected from a single resident's home. This may reduce how well the findings generalise to different homes. Secondly, there is some imbalance due to some locations being used less, such as staircase areas. This is practically acceptable, since places like corridors and staircases are used for transitions from one place to another, but are imbalanced for model training. Future releases of the dataset will fix these limits by collecting data from more homes and more people, and by recording for longer periods. This will make the results stronger and more reliable.

8 Conclusion and Future Work

The Robot House IoT Dataset (RH-IoT-1) supports research in assistive robotics and hospital-at-home environments. It provides a privacy-preserving resource for room-level human presence detection using ambient sensors. The dataset was collected in the University of Hertfordshire's Robot House, a research facility equipped with robots and dedicated to studies in HRI. Unlike many existing datasets, RH-IoT-1 includes spatial awareness, which is critical for assistive robots that must localise residents and coordinate safe interactions. It also supports context-sensitive decision-making, allowing robots to offer assistance in socially appropriate ways.

Baseline analysis conducted on the dataset shows that accurate room-level presence detection is achievable using non-intrusive sensing, supporting the effectiveness of privacy-friendly sensing for robotics. By providing location information, RH-IoT-1 enhance robots' context-aware capabilities. Future work should expand data collection to more households and longer deployments. This would improve generalisation and robustness for real-world robotic assistance scenarios.

References

1. World Health Organization. *Towards Age-friendly Primary Health Care.* World Health Organization, 2004.
2. Yoonseo Park, Eun-Ji Kim, Sewon Park, and Munjae Lee. Digital health intervention effect on older adults with chronic diseases living alone: systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*, 27:e63168, 2025.
3. CASAS, Washington State University. CASAS Datasets. Accessed: 2026-02-28.
4. Thomas Haferkamp, Samantha Ing, George Weyer, and Lauren J Gleason. Home-based primary care (HBPC): aging in place in 2025. *Current Geriatrics Reports*, 14(1):3, 2025.
5. Matteo Luperto, Javier Monroy, Jennifer Renoux, Francesca Lunardini, Nicola Basilico, Maria Bulgheroni, Angelo Cangelosi, Matteo Cesari, Manuel Cid, Aladar Ianes, et al. Integrating social assistive robots, iot, virtual communities and smart objects to assist at-home independently living elders: the movecare project. *International Journal of Social Robotics*, 15(3):517–545, 2023.

6. U Yeong Choi, Patrick Patterson, and Norma Chinho. Reducing the Burdens of Paid Caregivers of Older Adults by Using Assistive Technology: A Scoping Review . *Western Journal of Nursing Research*, 46(4):315–326, 2024.
7. Phani Teja Singamaneni, Pilar Bachiller-Burgos, Luis J Manso, Anaís Garrell, Alberto Sanfeliu, Anne Spalanzani, and Rachid Alami. A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*, 43(10):1533–1572, 2024.
8. Muhammad Zakir Khan, Muhammad Farooq, Ahmad Taha, Adnan Qayyum, Fehaid Alqahtani, Adnan Nadeem Al Hassan, Kamran Arshad, Khaled Assaleh, Shuja Ansari, Muhammad Usman, et al. Indoor localization technologies for activity-assisted living: Opportunities, challenges, and future directions. *Advances in Computers*, 133:59–98, 2024.
9. S Krishnakumar, Kalyan Acharjya, Shivam Khurana, Rakesh Ranjan Swain, Sourav Rampal, et al. Sensor Fusion Techniques for Accurate Indoor Tracking in IoT-Based Smart Environments. In *2025 International Conference on Automation and Computation (AUTOCOM)*, pages 754–759. IEEE, 2025.
10. Jesus D Ceron, Diego M López, Felix Kluge, and Bjoern M Eskofier. Framework for simultaneous indoor localization, mapping, and human activity recognition in ambient assisted living scenarios. *Sensors*, 22(9):3364, 2022.
11. Andrew Chan, Joanne Cai, Linna Qian, Brendan Coutts, Steven Phan, Geoff Gregson, Michael Lipsett, and Adriana M Ríos Rincón. In-home positioning for remote home health monitoring in older adults: systematic review. *JMIR aging*, 7(1):e57320, 2024.
12. Fatma Gongor and Onder Tutsoy. On the Remarkable Advancement of Assistive Robotics in Human-Robot Interaction-Based Health-Care Applications: An Exploratory Overview of the Literature. *International Journal of Human–Computer Interaction*, 41(2):1502–1542, 2025.
13. TP Vishnu, Debjit Ray, Karthick Thiyagarajan, and Abhra Roy Chowdhury. Multimodal Sensing for Socially Compliant Safe Robot Navigation in Human-Aware Indoor Environments Using Group-Aware Pose Estimation and Modified RRT. *IEEE Sensors Journal*, 2025.
14. Rushikesh A Deshmukh, Meghana A Hasamnis, Madhusudan B Kulkarni, and Manish Bhaiyya. Advancing indoor positioning systems: innovations, challenges, and applications in mobile robotics. *Robotica*, pages 1–41, 2025.
15. Sehrish Rafique, Patrick Holthaus, Gu Fang, and Farshid Amirabdollahian. Interpretable Room-Level Human Presence Detection Using Ambient Sensors in Smart Homes. In *17th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2025)*, 2025.
16. Ariza-Colpas P Patricia, Vicario Enrico, Butt Aziz Shariq, De-la_Hoz-Franco Emiro, Piñeres-Melo Marlon Alberto, Oviedo-Carrascal Ana Isabel, Muhammad Imran Tariq, Johanna Karina García Restrepo, and Patara Fulvio. Machine learning applied to datasets of human activity recognition: Data analysis in health care. *Current Medical Imaging Reviews*, 19(1):46–64, 2023.
17. Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate Activity Recognition in a Home Setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 1–9, 2008.
18. Hande Alemdar, Halil Ertan, Ozlem Durmaz Incel, and Cem Ersoy. ARAS Human Activity Datasets in Multiple Homes with Multiple Residents. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 232–235. IEEE, 2013.

19. Moritz Tenorth, Jan Bandouch, and Michael Beetz. The tum kitchen data set of everyday manipulation activities for motion tracking and action recognition. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 1089–1096. IEEE, 2009.
20. Niloy Sikder and Abdullah-Al Nahid. KU-HAR: An open dataset for heterogeneous human activity recognition. *Pattern Recognition Letters*, 146:46–54, 2021.
21. Nirmalya Thakur and Chia Y Han. Multimodal Approaches for Indoor Localization for Ambient Assisted Living in Smart Homes. *Information*, 12(3):114, 2021.
22. Nieves Pavón-Pulido, Juan Antonio López-Riquelme, Joaquín Ferruz-Melero, Miguel Ángel Vega-Rodríguez, and Antonio José Barrios-León. A service robot for monitoring elderly people in the context of Ambient Assisted Living. *Journal of Ambient Intelligence and Smart Environments*, 6(6):595–621, 2014.
23. Matteo Luperto, Javier Monroy, J Raul Ruiz-Sarmiento, Francisco-Angel Moreno, Nicola Basilico, Javier Gonzalez-Jimenez, and N Alberto Borghese. Towards Long-Term Deployment of a Mobile Robot for at-Home Ambient Assisted Living of the Elderly. In *2019 European Conference on Mobile Robots (ECMR)*, pages 1–6. IEEE, 2019.
24. Shigeo Tanabe, Eiichi Saitoh, Soichiro Koyama, Kei Kiyono, Tsuyoshi Tatemoto, Nobuhiro Kumazawa, Hitoshi Kagaya, Yohei Otaka, Masahiko Mukaino, Akira Tsuzuki, et al. Designing a robotic smart home for everyone, especially the elderly and people with disabilities. *Fujita medical journal*, 5(2):31–35, 2019.
25. Fahd Abuhoureyah, Wong Yan Chiew, Ahmad Sadhiqin Bin Mohd Isira, and Mohammed Al-Andoli. A Study of Bluetooth Propagation Using Accurate Indoor Location Mapping. *IET Wireless Sensor Systems*, 13(5):163–177, 2023.
26. Jianhua Liu, Zhijie Yang, Sisi Zlatanova, Songnian Li, and Bing Yu. Indoor Localization Methods for Smartphones with Multi-Source Sensors Fusion: Tasks, Challenges, Strategies, and Perspectives. *Sensors*, 25(6):1806, 2025.
27. O. Kerdjidj, Y. Himeur, S. S. Sohail, A. Amira, F. Fadli, S. Attala, W. Mansoor, A. Copiaco, A. Gawanmeh, S. Miniaoui, et al. Uncovering the potential of indoor localization: Role of deep and transfer learning. *IEEE Access*, 2024.
28. Daniela Micucci, Marco Mobilio, and Paolo Napoletano. UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones. *Applied Sciences*, 7(10):1101, 2017.
29. Patrick Holthaus, Trenton Schulz, Lewis Riches, Claudia-Andreea Badescu, and Farshid Amirabdollahian. ZTL: Lightweight Communication Patterns for HRI. 2026.