Simposio CEA de Robótica, Bioingeniería, Visión por Computador y Automática Marina 2026

An efficient hybrid HAR architecture for robust elderly AAL monitoring

Warnants, I.[a,*], Diallo, M.[b], Ortiz, F.J.[a], Garrigos, F.J.[c], Vera, J.A.[a]

[a]*Dpto. Ingeniería Eléctrica, Automática Industrial y Tecnología Electrónica, Universidad Politécnica de Cartagena member of European University of Technology EUT+, Campus Muralla del Mar, s/n, 30202 Cartagena, España.*
[b]*Networks and Telecommunications Department, University of Technology of Troyes (UTT), Troyes, France.*
[c]*Departamento de Electrónica, Tecnología de Computadoras y Proyectos, Universidad Politécnica de Cartagena member of European University of Technology EUT+, Plaza del Hospital, 1, Cartagena 30202, Spain*

Resumen

El reconocimiento de actividades humanas (HAR) es clave en sistemas de vida asistida por el entorno (AAL) para apoyar la vida independiente de personas mayores con robots sociales. Usando sensores ambientales no invasivos, el sistema garantiza privacidad sin cámaras intrusivas. Los modelos tradicionales son computacionalmente pesados y requieren un costoso etiquetado manual. Este trabajo presenta una arquitectura jerárquica eficiente de despliegue rápido. Nuestra metodología evita la anotación manual mediante abstracción por reglas y generación de datos sintéticos. Estas características entrenan un modelo híbrido TCNN-LSTM ligero de doble resolución. Al integrar contextos a corto (2 minutos) y largo plazo (2 horas), el modelo captura acciones transitorias y estados persistentes. Las pruebas reales lograron 95,09% de precisión y un macro F1 de 0,9486; la transferencia sintético-real alcanzó 91,33% de precisión y un macro F1 de 0,8412. Este pipeline ofrece una solución escalable y robusta para la monitorización doméstica en tiempo real.

Palabras clave: Percepción y detección robótica, Aprendizaje automático y profundo para la identificación de sistemas, Fusión y minería de datos en el control, Robótica social y ética, Seguridad y privacidad en sistemas ciber-físico-humanos, Sistemas de IA y automatización centrados en el ser humano y agencia humana, Aprendizaje automático para el modelado y la predicción, Aprendizaje por refuerzo y aprendizaje profundo en el control.

Abstract

Human Activity Recognition (HAR) is an important component of Ambient Assisted Living (AAL) systems supporting independent living for older adults with a social robot. Using non-invasive ambient sensors, the system ensures privacy without intrusive cameras. Traditional HAR models are often computationally heavy and rely on expensive manual labelling. This work introduces an efficient hierarchical architecture for rapid deployment. Our methodology bypasses manual annotation through rule-based abstraction and synthetic data generation. These features train a lightweight, dual-resolution hybrid TCNN-LSTM model. By integrating short-term (2-minute) and long-term (2-hour) contexts, the model captures transient actions and persistent behavioural states. Real-world tests achieved 95.09% accuracy and a 0.9486 macro F1-score; synthetic-to-real transfer reached 91.33% accuracy and a 0.8412 macro F1-score. This pipeline offers a scalable, robust solution for real-time domestic monitoring.

Keywords: Robot perception and sensing, Machine and deep learning for system identification, Data fusion and mining in control, Social robotics and ethics, Security and privacy in CPHS, Human-centric automation/AI Systems and human agency, Machine learning for modelling and prediction, Reinforcement learning and deep learning in control.

1. Introduction

The increasing elderly population and the high cost of professional homecare create an urgent need for Ambient Assisted Living (AAL) systems (Perez et al., 2023). The ADDIM project (ADDIM Project, 2025) addresses this by using wearables, ambient sensors, and social robotics designed to monitor user routines and improve environmental perception (Liu et al., 2024).

This research proposes a hybrid Human Activity Recognition (HAR) approach where rules handle the initial data abstraction, providing a reliable input for deep learning layers to inter-

[*]Autor para correspondencia: ismael.warnants@edu.upct.es

pret broader behavioural trends. By using a rule-based layer to map sensor events to simple activities, we eliminate the need for manual labelling. These features then feed into a hybrid TCNN-LSTM model that captures both short and long-term context. While our approach prioritises solo occupancy monitoring, it explicitly incorporates a CA_Visitor activity to identify periods when visitors or caregivers are present. This ensures that co-occupancy events are isolated, preventing others' sensor triggers from degrading the recognition of the primary user's independent routines while maintaining computational efficiency for edge deployment.

2. Related work

HAR research typically balances privacy and accuracy across vision, wearable, and environmental sensors. While vision and wearables provide rich data, environmental sensors are often preferred for long-term monitoring and abnormal behaviour detection due to their non-intrusive nature (Ankalaki et al., 2024; Wang et al., 2023). Modelling these event-based streams usually requires windowing strategies to capture temporal dependencies (Ankalaki et al., 2024; Min et al., 2020).

Current detection strategies range from interpretable rule-based systems to deep learning architectures. While rules offer stability, they fail to generalise across noisy signals or complex routines. Deep learning models, such as TCNNs and LSTMs, excel at capturing local patterns and sequential dependencies but often require massive labelled datasets and high-end hardware (Ramos et al., 2022). Expanding on architectural implementations, recent studies have emphasised the need for robust software patterns in IoT-based AAL systems (Mendes et al., 2023). Furthermore, recent work has explored LLM-based approaches to make recognition layout-agnostic (Thukral et al., 2024) or to improve clinical decision-making in robot-assisted environments (Martínez-Pascual et al., 2025). However, these advanced models often come at a very high computational cost.

This paper bridges the gap between deterministic rules and deep learning. By using rules for simple activity abstraction, we provide a semantic foundation for a neural network to model complex behaviours. We also utilise synthetic data generation to bypass the need for manual labelling of real-world datasets, which avoids the high hardware and data requirements of typical end-to-end pipelines.

3. Methodology

3.1. Hierarchical Activity Abstraction

The proposed HAR system uses a three-level hierarchy to separate physical sensor events from high-level behavioural understanding. As illustrated in Figure 1, we start by using a "Rule-Based Mapping" to translate "Raw Sensor Events" into "Simple Activities" at the first level (shown on the left). This allows us to perform a form of deterministic feature engineering. This step effectively filters sensor noise and reduces the complexity of the data before it reaches the neural network, which helps prevent overfitting. The first layer starts by capturing the sensor data to detect these foundational activities. From these simple activities, we can detect patterns using a "Machine Learning / Deep Learning Model" (shown in the centre) to understand "Complex Activities" like sleeping or cooking (shown on the right). We start by describing the sensors to correlate them with simple activities.

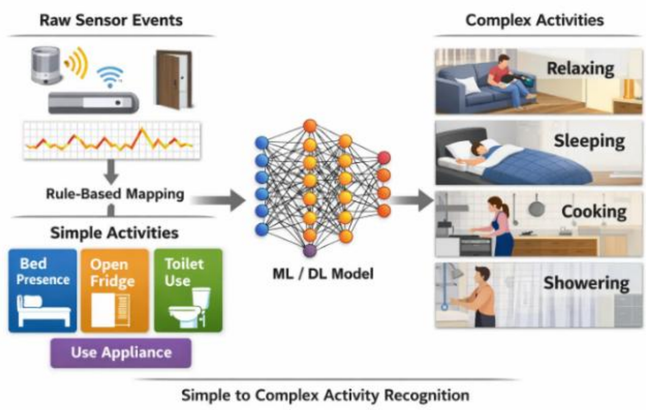

Figure 1: Structured pipeline for Human Activity Recognition (HAR).

3.1.1. Sensor data to simple activities

Within the ADDIM project, the system has been deployed in two real-world scenarios. The first installation, as shown in Figure 2, is located in an adapted apartment at the Fundación Poncemar in Lorca, where elderly users interact with the system under the supervision of professional therapists. The second deployment is in the private home of a 93-year-old woman with reduced mobility, yet maintaining functional autonomy, who lives alone in Cartagena and receives morning assistance from a caregiver. Both the adapted apartment and the private residence are equipped with a comprehensive suite of ambient sensors designed to capture human behaviour through multiple modalities.

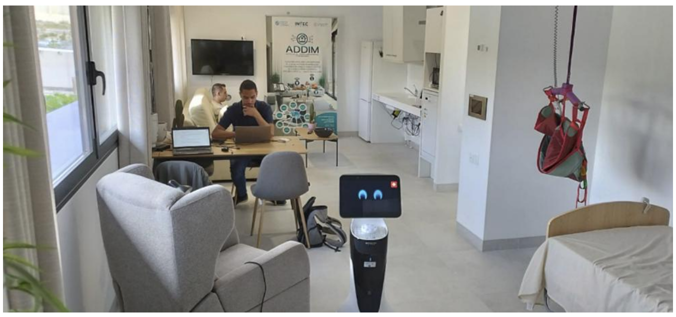

Figure 2: Adapted apartment and Orion Star Mini social robot at the facilities of the Poncemar Foundation in Lorca.

Movement and room occupancy are tracked using Passive Infrared (PIR) motion sensors, while pressure sensors in the bed and seating areas identify specific states such as resting or sleeping. Object interaction is monitored via contact sensors on doors and appliances, supplemented by vibration sensors that detect fine-grained furniture manipulation. The context is further enriched by power consumption sensors used to identify activities like cooking or watching TV, alongside temperature and humidity sensors that provide additional cues regarding the quality of life of the user.

These sensors are mapped to simple activities via rules after passing through a filtering layer. This layer employs a type of debouncing to filter noise and stabilise events during transitions, such as user movement in a seat. Deterministic rules then identify specific states; for instance, *A_UserSeated* triggers if any seat sensor is active, while *A_IdleBathroom* persists if no movement occurs for 10 minutes but the user has not exited the room. To illustrate the mechanics of this first layer, Listing 1 provides the corresponding pseudocode for the aforementioned examples, highlighting the deterministic nature of the rule-based abstraction.

```
// Example 1: Direct sensor rule
IF Chair_Sensor is ACTIVE OR Sofa_Sensor is
    ACTIVE:
    Trigger Activity "A_UserSeated"

// Example 2: Combined timer and location rule
IF User_Location is "Bathroom":
    IF Time_Since_Last_Motion >= 10 minutes:
        Trigger Activity "A_IdleBathroom"
```

Listing 1: Pseudocode showing the logic for simple activity rules.

After saving this first layer of information, it can be used to determine complex activities by looking at the duration and frequency of these simple activities. For this, we use deep learning.

3.1.2. Simple activities to complex activities using synthetic data

The model classifies behavioural patterns into seven high-level complex activities. Central to this is the CA_Visitor activity, which is designed to flag periods where sensor triggers cannot be reliably attributed to the primary user due to the presence of a second person, such as a caregiver. By categorising these moments separately, the system preserves the integrity of the data used for solo-user behaviours like sleeping or relaxing.

To train these models without the burden of manual labelling, we utilise a synthetic data generation process. A typical one-week routine created with the user serves as the foundation for a top-down stochastic generation process. High-level activities are scheduled within probabilistic time windows that determine occurrence and duration. Each complex activity is then populated via a stochastic action profile of simple actions. For instance, a cooking activity profile might include using the fridge, opening drawers, or activating the stove. This approach enables the system to provide monitoring and detection from the first day of installation, bypassing the traditional requirement of waiting months for data collection and manual annotation. Furthermore, it allows for a model highly optimised to the user's specific layout and habits, resulting in greater computational efficiency compared to oversized, general-purpose models.

To mitigate "perfect case" bias and ensure the model handles non-ideal scenarios, we introduce stochastic jitter to temporal boundaries and randomise the inclusion and sequence of these specific actions. This approach purposefully incorporates variability and simulated sensor failures, such as random dropouts, allowing the model to generalise beyond rigid, correct patterns to the noisy distributions of real-world behaviour.

3.2. Model creation

To create the model, we started with a basic architecture and evaluated its ability to interpret several weeks of data. The model was iteratively expanded to adapt to more complex synthetic data, ensuring it could understand the correlation between simple activity patterns and complex behaviours while avoiding overfitting. Features are constructed by combining binary vectors of simple activities with cyclical time-of-day features using sine transformations.

To capture both sudden transitions and persistent states, the data is segmented into two parallel sliding windows: a 2-minute short-term window and a 2-hour long-term context window. The 2-minute window captures quick actions like opening a door, while the 2-hour window is designed to cover the natural duration of typical household routines such as cooking, resting, or social visits. This choice ensures the temporal context aligns with the actual time these activities take to complete.

The hybrid TCNN-LSTM model uses Temporal Convolutional Neural Networks with dilated convolutions to identify local patterns. These dilated layers allow the model to look across longer periods of time without significantly increasing the number of parameters. This results in a system that can model long-range dependencies while remaining computationally efficient enough for real-time use. Figure 3 shows the layout of the model layers.

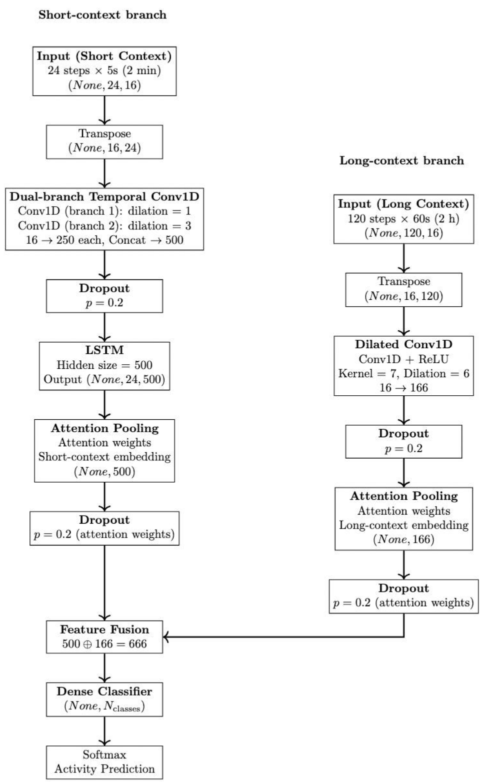

Figure 3: Hybrid TCNN-LSTM architecture with dual temporal resolution.

Training was performed on 20 labelled real days of ambient sensor data and completed within 4 minutes on an Apple M2-based MacBook Pro using integrated GPU acceleration. The fact that the model can be trained in minutes on a standard laptop shows its suitability for practical use. This low computational requirement is essential for running the system on embedded computers with limited hardware, allowing for real-time processing directly at the edge.

3.3. Model evaluation

The model's performance was evaluated using standard classification metrics including accuracy, precision, and recall. To address significant class imbalance, such as long-duration sleeping periods overshadowing shorter activities, the macro F1-score was utilised as the primary metric to ensure each activity class was given equal importance. Additionally, a class-weighted loss function was implemented during training to further penalise the model for errors on rare activities and ensure balanced learning.

3.4. Integration of HAR pipeline into the ADDIM system

The HAR pipeline functions as the perception layer for the ADDIM system. Sensor data is retrieved via MQTT and stored in InfluxDB. The pipeline performs real-time analysis and publishes detected activities back to the MQTT broker.

This allows a social robot to provide context-aware assistance, such as medication reminders, while avoiding interruptions during rest periods. Caregivers are notified of anomalies via real-time Telegram alerts and can access a Streamlit dashboard to review detailed activity metrics (Figure 4), as well as sleep and bathroom visit patterns (Figure 5). This approach balances automated monitoring with human oversight.

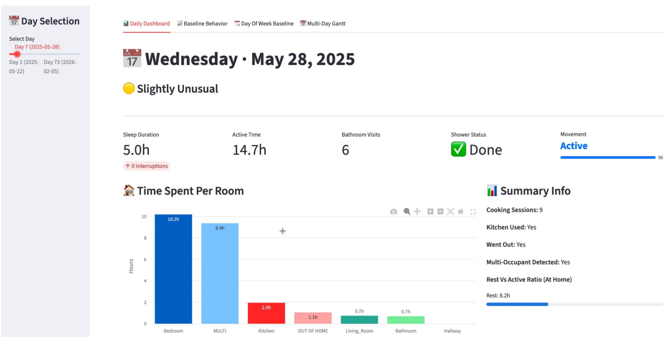

Figure 4: Dashboard for caregivers showing daily activity metrics.

4. Results & Discussion

4.1. Activities detected

With the structured pipeline, the system detects a total of 22 activities, consisting of 15 simple activities and 7 complex activities. Simple activities are identified through deterministic rules and grouped into categories to facilitate recognition as shown in Table 1.

Table 1: The activity categories with their simple activities

Category name	Simple activities
Presence and State	A_BedPresence, A_SeatedChair, A_SeatedSofa
Sanitation	A_WaterShower, A_ToiletUse, A_IdleBathroom
Domestic Interaction	A_KitchenAppliance, A_MicrowaveUse
	A_TakingThingsFridge, A_TakingThingsCupboard
Location Cues	Loc_Bedroom, Loc_Bathroom, Loc_Kitchen
	Loc_LivingRoom, Loc_Hallway

These simple activities provide the semantic features for the deep learning layer to predict high-level complex activities: CA_Sleeping, CA_Relaxing, CA_Cooking, CA_Visitor, CA_OutHome, CA_BathroomActivity, and CA_Shower.

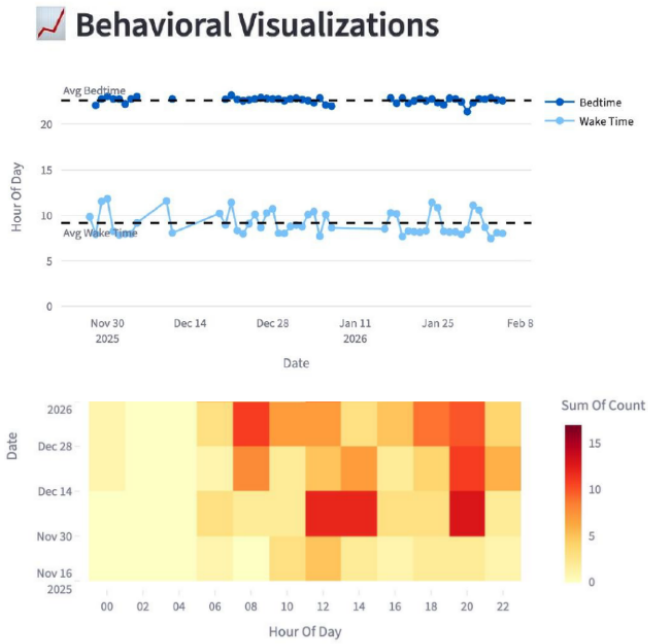

Figure 5: Visualisations of sleep patterns (top) and bathroom visit patterns (bottom).

4.2. Performance evaluation

Synthetic data were generated to replicate the real elderly user's home environment in Cartagena. To evaluate their realism and transferability, the model was trained exclusively on synthetic days and subsequently evaluated on real-world data under two validation configurations.

The real-world test involved a user with reduced mobility who sleeps on a living room sofa. This creates overlap between 'Relaxing' and 'Sleeping' signatures. Caregiver visits and social interactions were explicitly classified under the CA_Visitor activity category to prevent visitor movements from being misidentified as the primary user's actions. Since the system is optimised for solo monitoring, these co-occupancy periods were excluded from the evaluation of individual activities like cooking and showering. While this exclusion maintains the accuracy of the solo-activity profiles, it contributed to the data scarcity observed for those specific classes in the current 20-day test set.

No explicit background activity class was introduced in the current dataset. Observed inactivity gaps were minimal and generally corresponded to transition periods within complex activities rather than independent behavioural states. Therefore, a dedicated background class was not considered necessary at this stage. Figure 6 shows the test per-class accuracy and F1-score across detected Complex Activities.

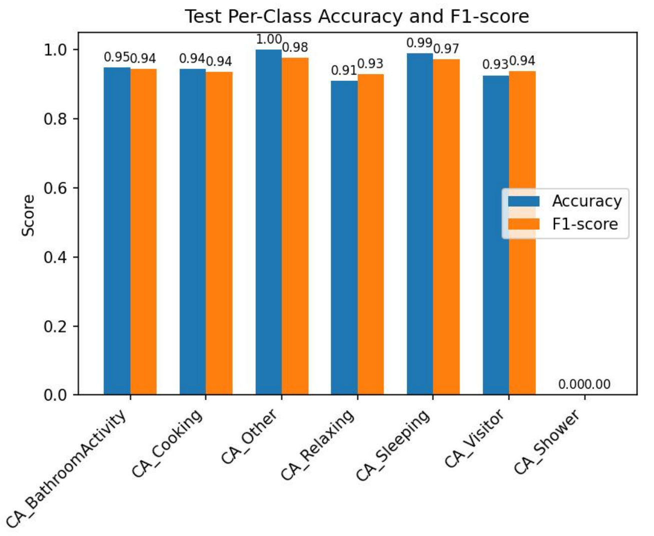

Figure 6: Test set accuracy and F1-score across Complex Activity (CA) classes. Results for CA_Shower are zeroed because of data scarcity in the initial testing phase.

Validation results achieved 96.00% accuracy, while test performance reached 95.09%. As mentioned before, the model faced challenges with shorter complex activities. Specifically, the lower recall in CA_Cooking (0.94 F1) and CA_BathroomActivity (0.94 F1) was caused by sensor ambiguity during CA_Visitor activities where a caregiver was present. Furthermore, the CA_Shower activity yielded a zero evaluation value due to the data scarcity mentioned previously, despite the simple activity A_WaterShower trigger being correctly identified at the rule layer. The test dataset currently contains 20 days of real data, which will increase as the monitoring of the user continues.

4.2.1. Synthetic-to-Real Transfer Experiment

To validate synthetic data generation, two experiments were conducted using exclusively synthetic data for training. In the first setup, the model was trained on 10 synthetic days, validated on two real days, and tested on two additional real days. This configuration achieved a test accuracy of 91.33% and a macro F1-score of 0.8412, indicating strong transfer from simulated to real ambient sensor data.

In the second setup, the model was trained and validated exclusively on synthetic data before being evaluated on real test days. While validation performance on synthetic data was perfect (Accuracy = 100%, Macro F1 = 1.00), test performance on real data decreased to 90.59% accuracy and a 0.8481 macro F1-score, highlighting the remaining gap between simulated and real activity distributions.

Although conducted on a limited number of real days and excluding CA_Visitor occupancy scenarios due to generation complexity, these results suggest that synthetic data can serve as an effective pre-training strategy. This opens the possibility of initialising activity recognition models prior to sensor deployment in new homes, reducing reliance on extensive labelled real-world data.

Rule-Based Baseline

Prior iterations of the ADDIM system relied solely on deterministic rules. While interpretable, this approach proved insufficient for complex and ambiguous activities. It struggled particularly with overlapping behaviours and temporal dependencies, leading to unstable recognition performance. These limitations motivated the integration of deep learning for higher-level activity modelling while preserving rule-based semantic abstraction.

End-to-End Pipelines

The SDHAR-HOME pipeline serves as a benchmark for high-accuracy, end-to-end deep learning (Ramos et al., 2022). Table 2 shows a comparison of some features between the SDHAR-HOME pipeline and the pipeline proposed in this paper. The SDHAR-HOME model requires two months of data, multimodal sensing including wearables, and high-end hardware like dual RTX 3090 GPUs, leading to an 18-hour training period.

In contrast, our structured approach achieves a superior test accuracy of 95.09% and a 0.9486 macro F1-score using only 20 days of labelled activity data and a consumer-grade laptop for a few minutes of training. While the SDHAR-HOME model requires a two-month data collection period before becoming operational, our hierarchical structure is ready for deployment on day one. This flexibility ensures that the system provides immediate value to the elderly user and their caregivers without an extensive or intrusive calibration phase. By leveraging synthetic data, the model is highly optimised to the user's specific habits and layout, making it significantly more efficient than oversized, general-purpose models. However, it is important to note that the proposed pipeline only detects one person and cannot follow the user when there are multiple people present, in contrast to the SDHAR-HOME pipeline that is able to detect two people.

Table 2: Feature comparison of the SDHAR-HOME pipeline with the proposed pipeline

Feature	SDHAR-HOME	Proposed Structure
Dataset duration	2 months	20 labelled days
Residents	2	1
Activity classes	18	Hierarchical (22 total)
Sensor modalities	Wearables, ambient	Ambient sensors only
Architecture	LSTM end-to-end	TCNN-LSTM
Test Accuracy	88-90%	95.09%
Macro F1-score	User 1: 0.88, User 2: 0.82	0.9486
Hardware used	2x RTX 3090 GPUs	Apple M2 (integrated GPU)
Training time	18 hours	Few minutes

LLM-Based Abstraction pipeline (TDOST)

The TDOST framework utilises natural language descriptions and large language models for layout-agnostic recognition. While TDOST achieves competitive results (90-96% F1), it imposes significant computational burdens due to the use of heavy LLM encoders. Our TCNN-LSTM pipeline provides similar accuracy (95.09%) but remains far more lightweight and suitable for local edge deployment.

4.3. Limitations

Occupancy Ambiguity

The reliance on non-intrusive ambient sensors makes it difficult to distinguish between the primary user and visitors like caregivers. Consequently, these actions are labelled as the

CA_Visitor activity, which can overshadow specific behaviours like cooking. The system prioritises solo monitoring as visitors provide immediate support and sensor triggers cannot be reliably attributed to one person.

Behavioural Overlap

Physical challenges, such as the user sleeping on a living-room sofa due to limited mobility, create a notable overlap between ”Relaxing” and ”Sleeping” signatures, complicating class separation.

Data Scarcity

The short 20-day real-world testing window provided insufficient instances of rare activities such as CA_Shower, and this activity was not present in the test split. Consequently, no evaluation metrics could be computed for this category, and the reported zero value reflects the absence of ground-truth samples rather than a model recognition failure.

Partial Observability of Ambient Sensors

Unlike cameras, ambient sensors provide only ”clues” about behaviour. They are prone to false positives (triggers without meaningful activity) and false negatives (missed triggers when an activity occurs or when a sensor fails).

5. Conclusion

This work introduces a lightweight, hierarchical HAR architecture specifically designed for robust monitoring in elderly AAL environments. By integrating a deterministic rule-based abstraction layer with a dual-resolution TCNN-LSTM model, we effectively bypass the need for intrusive cameras and burdensome manual annotation.

Real-world tests achieved 95.09% accuracy and a 0.9486 macro F1-score using minimal training data. Synthetic-to-real transfer tests reached 91.33% accuracy and a 0.8412 macro F1-score, validating the pre-training strategy. The dual-context strategy captures sudden transitions and persistent states, enabling social robots to deliver proactive assistance. Ultimately, this scalable, privacy-preserving pipeline deploys rapidly on low-cost hardware to support independent living.

6. Future Work

Future work will focus on extending the data collection period beyond the initial 20-day window to increase dataset diversity and improve the robustness of rare activity detection. Expanding the real-world dataset will allow more stable evaluation and contribute to improved dashboard accuracy.

A key priority is the automation of the labelling process to reduce manual intervention and simplify the deployment process. This includes refining the structured abstraction layer to generate more reliable semi-automatic annotations.

In parallel, we will refine the activity monitoring dashboard in collaboration with psychologists and psychotherapists for testing under real-world conditions. We aim to improve the clarity of the interface by adding temporal summaries and anomaly detection, which will make the information much easier for caregivers and families to interpret. These developments are intended to increase the stability and practical reliability of the system for long-term home monitoring.

Acknowledgements

The research presented in this paper has been developed within the framework of the ADDIM project (2023–2026), a Public–Private Partnership Project (CPP2022_009649) funded by MICIU/AEI /10.13039/501100011033 and the European Union NextGenerationEU/PRTR. The project is carried out in collaboration with IntecRobots and the Universidad de Murcia as consortium partners.

The authors also acknowledge Fundación Poncemar (Lorca, Spain) and the UPCT–Poncemar Chair for facilitating the use of the adapted apartment for system deployment and evaluation, and for the involvement of the professional therapists who supervised the experimental activities.

References

ADDIM Project, 2025. Addim project. https://www.upct.es/noticias/2025-05-30-testan-en-casa-de-una-nonagenaria-el-sistema-de-apoyo-digital-domiciliario-integral-para-mayores, accessed: 2026-02-26.

Ankalaki, S., et al., 2024. Simple to complex, single to concurrent sensor-based human activity recognition perception. ArXiv.

Liu, X., Huang, C., Zhu, H., Wang, Z., Li, J., Cangelosi, A., 2024. State-of-the-art elderly service robot environmental perception, compliance control, intention recognition, and research challenges. IEEE Systems, Man, and Cybernetics Magazine. DOI: 10.1109/MSMC.2023.3238855

Martínez-Pascual, D., Rubira-Úbeda, P., Reig-Lozano, A., Martín-Batanero, R., Lledó, L. D., García-Aracil, N., 2025. Diseño de un sistema basado en llms para mejorar la toma de decisiones en rehabilitación asistida por robots. In: Simposio CEA de Robótica, Bioingeniería, Visión Artificial y Automática Marina 2025.

Mendes, L. F., Aguilar, P. A. C., Bezerra, C. I. M., 2023. Software architecture for iot-based indoor positioning systems for ambient assisted living. In: 2023 IEEE 20th International Conference on Software Architecture (ICSA). DOI: 10.1109/ICSA56044.2023.00017

Min, W., et al., 2020. Learning-based segmentation of activities of daily living in smart homes. IEEE Transactions on Automation Science and Engineering.

Perez, A. J., Siddiqui, F., Zeadally, S., Lane, D., 2023. A review of iot systems to enable independence for the elderly and disabled individuals. Internet of Things 21, 100653. DOI: 10.1016/j.iot.2022.100653

Ramos, R. G., Domingo, J. D., Zalama, E., et al., 2022. Sdhar-home: A sensor dataset for human activity recognition at home. Sensors 22, 8109. DOI: 10.3390/s22208109

Thukral, M., Dhekane, S. G., Hiremath, S. K., et al., 2024. Layout-agnostic human activity recognition in smart homes through textual descriptions of sensor triggers (tdost). ArXiv.

Wang, Y., Wang, X., Arifoglu, D., Lu, C., Bouchachia, A., Geng, Y., Zheng, G., 2023. A survey on ambient sensor-based abnormal behaviour detection for elderly people in healthcare. Electronics 12 (7), 1539. DOI: 10.3390/electronics12071539