2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

A ROS-Based Multi-Modal Architecture for Fall Detection and Response with a Social Robot

Kavyan Zoughalian[1], Imene Tarakli[1], Aung Htet[1], Joshua Bamforth[1], Alejandro Jiménez-Rodriguez[1], Jims Marchang[1], Alessandro Di Nuovo[1]

***Abstract*— Falls are a leading cause of injury in older adults, requiring detection systems that are both sensitive and reliable. We present a multi-modal robotic framework that integrates wearable sensing, vision-based verification, and dialogue-driven assessment. A smartwatch streams inertial data, with thresholds tuned through pilot testing to maximise fall sensitivity. Vision verification is performed using a fine-tuned YOLOv11 model, while Whisper ASR and a lightweight GPT-based classifier enable simple verbal checks of user responsiveness. Our tuned thresholds outperformed published baselines (F1 = 0.857), and the vision module achieved strong accuracy (mAP@0.5 = 0.827). In integrated trials, the system reached a 90.6% success rate with a mean end-to-end response time of 43.5 seconds. These results show that combining complementary modalities enhances robustness and moves socially assistive robots toward interactive fall response in real-world care.**

I. INTRODUCTION

Global demographic shifts are leading to an unprecedented rise in the proportion of older adults. By 2050, one in six people worldwide will be over the age of 65, with many expected to live well into their 80s and beyond [1]. Although longevity is a testament to advances in healthcare, it also brings increased vulnerability to age-related risks, most notably falls. Falls are the leading cause of injury-related hospitalisations and deaths among people aged 65 and older, with roughly 30% of community, dwelling older adults experiencing at least one fall annually [2], [3]. Such incidents often have serious consequences, including fractures, loss of independence, and premature mortality. Moreover, prolonged time lying on the floor after a fall is strongly associated with worse outcomes, making timely detection and response critical [4].

Older adults consistently express a desire to maintain independence and remain in their own homes for as long as possible [5]. However, delivering continuous human supervision is often impractical due to workforce shortages, caregiver fatigue, and the financial burden of 24/7 monitoring [6]. Assistive robotics has emerged as a promising means to support ageing-in-place by providing timely assistance and health monitoring [7]. Yet, relying solely on robot-based monitoring can be intrusive, particularly in private spaces such as bathrooms, raising privacy concerns and potentially impacting user acceptance [8].

A practical alternative is to deploy multi-modal fall detection systems in which a wearable device continuously monitors the user's motion and only summons the robot when a potential fall is detected. Smartwatches, in particular, are an attractive option: they are unobtrusive, already widely adopted, and can host effective threshold-based fall detection algorithms with high specificity. In such a system, the wearable serves as the primary sensor, triggering a mobile robot to navigate to the user's location, verify the incident via vision, and interact verbally to assess severity before deciding whether to escalate.

In this paper, we propose a ROS-based multi-modal architecture, as seen in Figure 1, that integrates heuristic smartwatch fall detection, autonomous robot navigation, vision-based verification, and dialogue-based severity assessment. Our approach aims to minimise false alarms, reduce unnecessary intrusions, and ensure rapid, context-sensitive assistance. Unlike prior work that focuses on a single sensing modality or omits real-time integration, our framework enables an end-to-end, on-demand robotic response triggered by wearable sensing, balancing privacy, responsiveness, and reliability.

Our main contributions are:

1) The design and implementation of a ROS architecture integrating wearable-triggered fall detection with robotic navigation, visual confirmation, and conversational assessment.
2) A modular communication pipeline enabling low-latency, event-driven wearable–robot interaction.
3) An experimental evaluation of the system's detection accuracy, component performance, and end-to-end re-

Fig. 1. ARI social robot approaching a participant simulating a fall during system evaluation.

*This work was not supported by any organisation

[1]Department of Computing, Advanced Wellbeing Research Centre (AWRC), Sheffield Hallam University, UK. k.zoughalian@shu.ac.uk

sponse time in realistic scenarios.

II. RELATED WORK

Fall detection approaches can be broadly categorised into wearable, vision-based, and ambient sensing systems.

Wearable systems rely on inertial measurement units (IMUs) embedded in devices such as smartwatches. Threshold-based heuristics remain popular due to their low latency and computational simplicity, e.g., PIPTO [9] and BLE-based detectors [10], which monitor impact and post-fall inactivity. While these systems achieve high specificity, they often suffer from false alarms and lack automatic verification on site.

Vision-based systems, such as those using pose-estimation or deep learning, avoid the need for wearables and provide contextual awareness. For example, Pfyffer et al. [11] integrated a YOLOv8 detector into the *tēmi* robot to automatically identify and report falls in healthcare environments. Despite promising results, vision-only systems face challenges of line-of-sight, occlusion, and intrusiveness in private spaces.

Hybrid wearable–robot systems aim to balance privacy and coverage by triggering a robot only when a wearable signals a fall. Raeve et al. [10] demonstrated a BLE-based system that could alert caregivers, while other prototypes combine wearable-triggered navigation with onboard vision for confirmation [12]. However, most of these studies stop at component-level accuracy or vision confirmation, without addressing dialogue-based severity assessment or fully describing the integration of ROS-based pipelines.

Our work extends this literature by presenting a ROS-based, end-to-end framework where a heuristic smartwatch detector triggers robot navigation, vision-based verification, and dialogue-based severity assessment. Unlike prior efforts, we provide a quantitative evaluation not only of individual modules but also of full-system latency and failure modes, thereby addressing gaps in reporting integration and response times, under real participant conditions.

III. SYSTEM ARCHITECTURE

We propose a unified assistive robotics framework in the Robot Operating System (ROS) that integrates multi-modal fall detection, heterogeneous sensing (smartwatch accelerometer and onboard RGB camera), autonomous navigation, and dialogue-based health assessment with one coherent control system deterministic over the robot's actions, as illustrated in Figure 2. Unlike prior works relying solely on wearable devices [10] or stationary vision systems [13], our approach employs a wearable-triggered pipeline. The robot is deployed only after a potential fall is detected, at which point it verifies the incident and assesses the user's condition. This architecture reduces false alarms and preserves privacy while ensuring timely assistance.

Figure 3 illustrates the main ROS nodes that compose the system:

1) **Social Node:** enables the robot's general companion behaviours, including casual conversation and presence in the environment outside emergency events.
2) **Watch Node:** continuously acquires tri-axial accelerometer data from the smartwatch and applies the heuristic fall-detection algorithm. When a potential fall is detected, it sends a fall_detected service request.
3) **Navigation Node:** manages the robot's movement, using the ROS navigation stack to sequentially visit predefined *Points of Interest* (POIs) until the user is located.
4) **Vision Node:** is triggered once the robot reaches the suspected fall location. It processes the RGB camera feed using a YOLO-based pose estimator to provide a binary fall confirmation.
5) **Assessment Node:** is activated when both wearable and vision detections indicate a fall. It conducts a verbal health check through Whisper ASR and intent classification, determining whether the user is responsive or in need of help.
6) **Controller Node:** coordinates the pipeline by orchestrating service requests and responses across all modules. It ensures the correct sequence from wearable-triggered detection, through navigation and vision verification, to verbal assessment and logging.

Inter-module communication is managed through ROS action nodes and sensing nodes via service level communication as illustrated in Figure 3, enabling modular integration and simplifying future upgrades or substitutions.

A. Wearable Fall Detection

The smartwatch module implements a *heuristic rule-based* algorithm inspired by prior works [3]. Let $a(t)$ denote the magnitude of the tri-axial acceleration vector:

$$a(t) = \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2} \tag{1}$$

Three tunable parameters (Figure 4) define the detection logic:

- **Impact threshold** (a_{impact}): minimum acceleration magnitude to signal a potential fall.
- **Rest threshold** (a_{rest}): maximum acceleration magnitude permitted during the post-impact rest period.
- **Rest window** (T_{rest}): time interval following the impact where $a(t) \leq a_{\text{rest}}$ must hold to confirm a fall.

A fall is detected if:

$$\max_{t \in W_{\text{impact}}} a(t) \geq a_{\text{impact}} \tag{2}$$

$$\max_{t \in [t_{\text{impact}},\, t_{\text{impact}} + T_{\text{rest}}]} a(t) \leq a_{\text{rest}} \tag{3}$$

where W_{impact} is a short sliding window for peak detection and t_{impact} is the time of maximum acceleration.

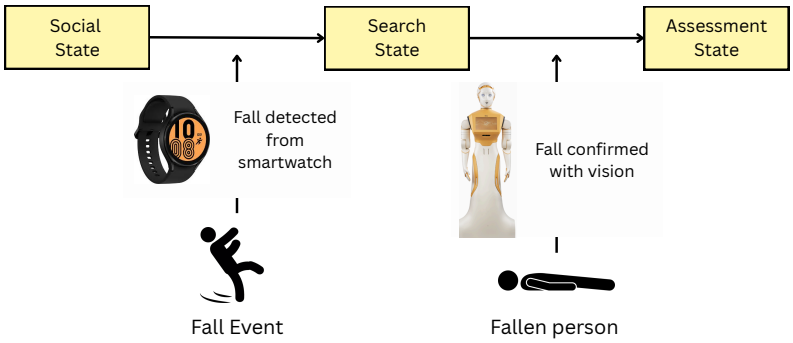

Fig. 2. System architecture of the fall detection. A smartwatch first detects a potential fall, which triggers the robot to enter the *Search State* and navigate towards the person. Once located, the fall is confirmed through onboard vision before the robot transitions to the *Assessment State* to evaluate the user's condition via dialogue.

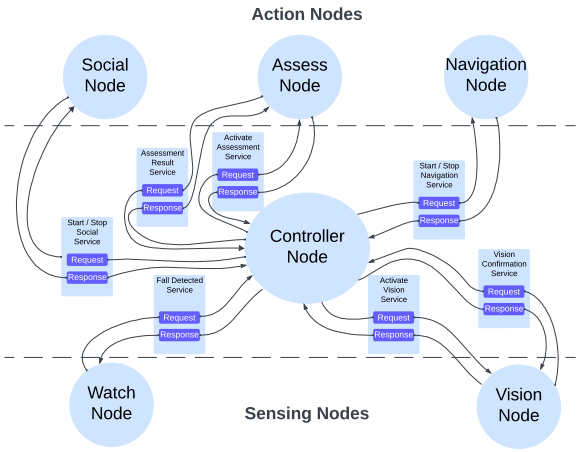

Fig. 3. ROS framework of the proposed system architecture.

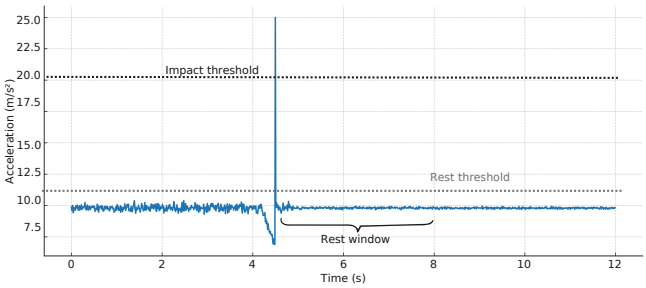

Fig. 4. Example acceleration magnitude profile during a simulated fall. The peak corresponds to the moment of impact, followed by a low-activity period (rest window) used to confirm the fall. The impact and rest thresholds define the detection criteria in the smartwatch.

B. Vision-Based Fall Verification

The vision-based verification module determines whether a person has fallen within the perimeter of the robot's torso-mounted RGB camera using a YOLOv11 object detection framework [14]. The model was fine-tuned on a publicly available fall detection dataset [2] to recognise characteristic postures associated with a fallen individual. As the robot navigates through the local environment through the use of POIs, image frames from the torso camera are processed in real time, and the model outputs a binary classification (fallen or not fallen) together with a confidence score. This acts as a second verification pipeline to reduce false alarms, while allowing the robot to get an approximate location of the fallen person within its environment. The confidence threshold provided by the model was empirically tuned to minimise false positives introduced by the model, balancing sensitivity to true falls with robustness against misclassifications. In this case, a confidence threshold of 0.7 has been introduced to the model for reducing misclassifications.

C. Dialogue-Based Condition Assessment

If both wearable and vision detections indicate a fall, the robot initiates a verbal health check by asking simple questions such as "Are you alright?" and "Do you need help?". This approach focuses on assessing basic responsiveness rather than performing a full fall evaluation. Speech recognition is handled using Whisper ASR [15], and responses are classified into two intents: responsive or unresponsive, using a lightweight GPT-based classifier [16]. The system currently logs these outcomes and is designed to allow future integration with caregiver alert systems.

IV. METHODOLOGY

A. Experimental setup

The proposed system was implemented and tested on a PAL Robotics ARI humanoid robot equipped with an RGB camera, microphone array, and speaker system. The robot was connected to an NVIDIA Jetson Orin Nano, running ROS 1, which hosted the perception, navigation, and dialogue nodes. The watch-based fall detection was deployed on a Samsung Galaxy Watch 5 Pro, which streamed tri-axial accelerometer data at 100 Hz to the ROS network via wireless communication. Experiments were conducted in a lab indoor environment arranged to simulate a typical apartment space. A 2D occupancy map was generated using SLAM and divided into four *Points of Interest* (POIs) corresponding

to key functional areas (e.g., living area, kitchen, bedroom, corridor) as seen in Figure 5. The navigation module was configured to sequentially visit these POIs when a potential fall was detected, enabling systematic coverage without requiring constant robot proximity to the user.

B. Study design

To evaluate the system, we recruited **eight adult participants** (6 male, 2 female, ages 24-31) from Sheffield Hallam University. Participation was voluntary, and all participants provided informed consent prior to the study.

Each participant was asked to perform four everyday scenarios designed to simulate falls or fall-like events:

1) **Forward fall:** intentionally falling forward onto a mat.
2) **Backward fall:** intentionally falling backward onto a mat.
3) **Kneeling:** lowering down to pick up an object from the floor.
4) **Lying on a couch:** sitting and reclining into a lying position.

The order of scenarios was randomised across participants to avoid sequence effects.

In addition to the physical scenarios, we simulated two dialogue outcomes when the assessment module was triggered:

- **Responsive:** participants answered verbally to indicate they were not injured (e.g., "I'm fine").
- **Unresponsive:** participants remained silent to simulate loss of responsiveness.

At the start of each trial, the participant wore the smartwatch while positioned in one of the designated POI on the map. Upon performing the scenario, the smartwatch fall detector streamed acceleration data to the ROS framework. If a potential fall was detected, the ARI robot navigated to the participant's location, performed vision-based verification, and initiated the dialogue-based assessment depending on the condition.

All system outputs and timestamps were logged for later analysis of detection accuracy, false alarms, and end-to-end response time.

C. Evaluation Metrics

We evaluated the system at both the *module level* (wearable, navigation, vision, dialogue) and the *end-to-end system level*. The following metrics were used:

1) Wearable Fall Detection: We evaluated the smartwatch-based fall detection using *precision*, *recall*, and *F1-score*.

2) Vision-Based Verification: The vision module was assessed using the same set of metrics (precision, recall, and F1-score) applied to the classification of RGB camera frames into *fall* and *no-fall* categories.

3) Navigation Performance: We measured the **localization time**, defined as the elapsed time from the initiation of robot navigation (after a wearable-triggered fall event) until the robot successfully reached the correct Point of Interest (POI) where the participant was located.

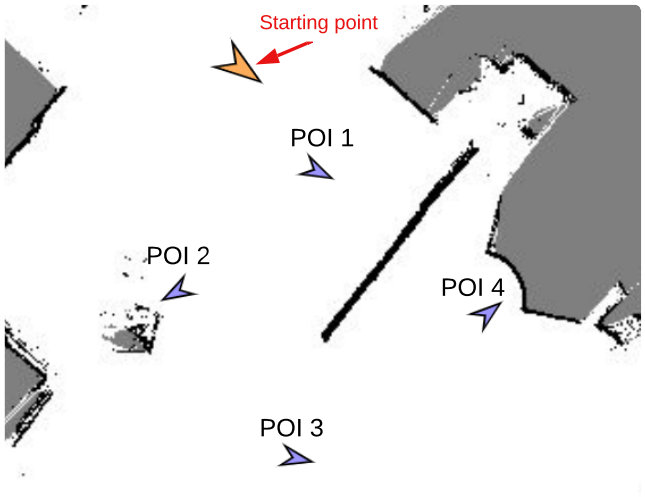

Fig. 5. Navigation map used during system evaluation. The robot sequentially searched four points of interest (POIs): POI1 was designated for kneeling activities, POI3 for simulated falls, and POI4 for lying down scenarios. These POIs allowed controlled testing of fall and non-fall cases under consistent spatial conditions.

4) Dialogue-Based Assessment: The dialogue module was evaluated using two complementary measures. First, we measured the **time-to-assessment**, defined as the duration from the initiation of verbal interaction to the point at which the system classified the participant as either *responsive* or *unresponsive*. Second, we assessed classification performance by computing precision, recall, and F1-score for the two intent categories, using the ground truth of the scripted participant responses as reference.

5) System-Level Performance: System-level evaluation focused on the effectiveness of the integrated pipeline. Two measures were considered:

- **Success rate:** proportion of trials in which the system completed the full pipeline, from fall detection to dialogue-based classification, without critical failure.
- **End-to-end latency:** average elapsed time from the fall event to the final dialogue-based classification.

V. Results and Discussion

A. Wearable Fall Detection Validation

We evaluated three threshold configurations for the smartwatch module (Table I): two taken directly from prior studies [9], [10] and one adapted through pilot testing to our 100 Hz data stream and experimental conditions. The latter was tuned to maximise sensitivity to true falls, even at the cost of occasional false positives, since missed falls are the most critical failure. This evaluation was conducted as a pre-study with five participants, each performing both fall (forward and backward) and non-fall (kneeling or lying down) scenarios, with metrics reported in terms of sensitivity, specificity, precision, and F1-score.

Table II shows that our tuned threshold configuration substantially outperformed both published baselines. While BLE offered moderate balance between precision and recall,

TABLE I

THRESHOLD VALUES

Threshold Type	PIPTO [9]	BLE-Fall [10]	Ours
Impact Threshold	30.0	24.5	20.0
Rest Threshold	9.8	7.85	11.0
Rest Window (ms)	850	1000	1000

and PIPTO achieved higher precision but at the cost of very low recall, our configuration reached an F1-score of 0.857, with recall close to 0.9. This indicates that it was not only more sensitive to true falls but also maintained high precision, avoiding excessive false alarms. The results support our design choice to prioritise sensitivity, as missing genuine falls poses a greater risk than tolerating occasional false positives.

TABLE II

PERFORMANCE COMPARISON OF FALL DETECTION METHODS.

Method	Precision	Recall	F1 Score
BLE	0.700	0.636	0.667
PIPTO	0.750	0.250	0.375
Ours	**0.818**	**0.900**	**0.857**

B. Vision Fall Validation

This subsection evaluates how the model performs in terms of its recall, precision and F1 score for different confidence score introduced by the model and how this performs in terms of its validate and test set provided by the dataset. At lower thresholds ($\leq$ 0.4), recall remains high but precision is comparatively lower, indicating more false positives. As the threshold increases, precision improves steadily, peaking around 0.85–0.9, while recall declines due to an increase in false negatives.

The chosen confidence threshold point of 0.7 (vertical dashed line) represents a balance between these metrics. At this threshold, precision is substantially improved over lower values while recall remains at an acceptable level, resulting in a competitive F1-score. This point prioritises reducing false positives, important in the context of wearable-triggered verification, while still maintaining sufficient sensitivity to true falls.

The YOLOv11 model, fine-tuned for fall detection with a confidence score of 0.7, achieved strong classification performance with an mAP@0.5 of 0.867 on the validation set and 0.827 on the test set (Table IV). This validation confirmed that a confidence threshold of 0.7 was suitable for integration into the full system evaluation.

TABLE III

YOLO MODEL PERFORMANCE ON VALIDATION AND TEST SETS-A

Split	Precision	Recall	F1
Validation	0.831	0.802	0.816
Test	0.808	0.753	0.780

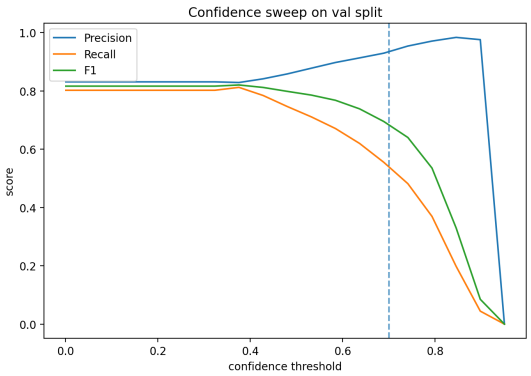

Fig. 6. The figure shows the effect of varying the YOLOv11 confidence for threshold on precision, recall, and F1-score for the validation split in fall detection.

TABLE IV

YOLO MODEL PERFORMANCE ON VALIDATION AND TEST SETS-B

Split	mAP@0.5:0.95	mAP@0.5	mAP@0.75
Validation	0.522	0.867	0.557
Test	0.510	0.827	0.567

C. System Evaluation

After validating the individual modules, we conducted a full-system evaluation to assess the integrated performance of the architecture with participants. In this stage, the smartwatch-based fall detection employed our tuned threshold configuration, as it demonstrated the best balance of sensitivity and precision during pre-study testing. The evaluation focused on end-to-end performance across detection, navigation, vision verification, and dialogue-based assessment, providing insight into how the system operates under realistic conditions rather than isolated module testing.

Table V presents the performance of each module and the overall system. The watch-based fall detection maintained strong sensitivity (recall = 0.938) with reasonable precision, while the vision module achieved perfect recall, confirming all true falls, and higher precision than the wearable alone. The assessment module consistently classified responsiveness correctly across all trials. Taken together, these modules yielded an overall system success rate of 90.6%. This demonstrates that the multi-modal architecture not only mitigates the limitations of individual components but also achieves robust end-to-end performance in realistic scenarios. Importantly, the complementary strengths of wearable-triggered detection and vision-based verification contributed to reducing false positives while maintaining high fall sensitivity.

Furthermore, Table VI presents the temporal performance of the system. On average, the robot required 27.0 seconds to locate the participant, with a relatively high variance (SD = 13.5 s). This comparatively short localisation time was influenced by the constrained size of the laboratory test room, where only four points of interest needed to be searched. Once the participant was found, the assessment module classified responsiveness in 16.5 seconds on average,

TABLE V

PERFORMANCE METRICS FOR WATCH, VISION, AND OVERALL SYSTEM SUCCESS RATE.

System	Precision	Recall	F1 Score
Watch	0.833	0.938	0.882
Vision	0.882	1.000	0.938
Assessment	1	1	1
System Success Rate		0.906	

showing much lower variability. The overall end-to-end response time was therefore 43.5 seconds, which is reasonable for a controlled indoor environment. While the assessment phase was consistent across trials, the localisation phase remains sensitive to environmental layout and initial robot position. In larger, more realistic deployments, localisation time would likely increase, suggesting that optimisations such as dynamic path planning or wearable–robot position fusion could further reduce overall response times.

TABLE VI

MEAN, STANDARD DEVIATION, AND 95% CONFIDENCE INTERVALS OF DURATIONS.

Metric	Search (s)	Assess (s)	Total (s)
Mean $\pm$ SD	27.03 $\pm$ 13.47	16.46 $\pm$ 3.83	43.49 $\pm$ 12.30
95% CI	[19.97, 34.09]	[14.45, 18.47]	[37.10, 49.88]

VI. CONCLUSIONS

This study presented an integrated, multi-modal framework for fall detection and assessment combining a smartwatch, robot navigation, vision-based verification, and dialogue-based responsiveness checks. Module-level evaluations showed that our tuned threshold configuration for wearable detection substantially outperformed existing baselines, while the fine-tuned YOLOv11 model achieved strong recall and precision at a balanced confidence threshold. At the system level, the architecture successfully integrated detection, navigation, and assessment, achieving a 90.6% overall success rate with a mean end-to-end response time of 43.5 seconds.

These findings highlight the value of combining wearable-triggered detection with robot-based verification, as the complementary strengths of the modules reduced false alarms while maintaining sensitivity to true falls. The dialogue-based assessment module further ensured that user responsiveness could be checked safely and non-intrusively.

Nonetheless, the study was constrained by the controlled laboratory environment, a small participant pool, and relatively short localisation times due to the limited size of the testing area. In real-world deployments, larger environments and greater behavioural variability will introduce new challenges.

Future work will focus on scaling evaluations to more diverse participants and home-like dimension environment, optimising localisation through dynamic planning or wearable–robot position fusion, and integrating the communication interface to enable caregiver alerts. Ultimately, the proposed framework provides a promising foundation for socially assistive robots capable of delivering safe, timely, and personalised fall response in real-world contexts.

ACKNOWLEDGMENT

The author would like to thank the members of the Smart Interactive Technologies Lab for their invaluable support, collaboration, and contributions throughout this work.

REFERENCES

[1] Office of the High Commissioner for Human Rights (OHCHR), "About the human rights of older persons," https://www.ohchr.org/en/special-procedures/ie-older-persons/about-human-rights-older-persons, n.d., accessed: 2025-08-16.

[2] Centers for Disease Control and Prevention, "Older adult falls data," https://www.cdc.gov/falls/data-research/index.html, 2024, accessed: 2025-08-16.

[3] M. Montero-Odasso, N. Van Der Velde, F. C. Martin, M. Petrovic, M. P. Tan, J. Ryg, S. Aguilar-Navarro, N. B. Alexander, C. Becker, H. Blain *et al.*, "World guidelines for falls prevention and management for older adults: a global initiative," *Age and ageing*, vol. 51, no. 9, p. afac205, 2022.

[4] E. Bayen, S. Nickels, G. Xiong, J. Jacquemot, R. Subramaniam, P. Agrawal, R. Hemraj, A. Bayen, B. L. Miller, and G. Netscher, "Reduction of time on the ground related to real-time video detection of falls in memory care facilities: observational study," *Journal of medical internet research*, vol. 23, no. 6, p. e17551, 2021.

[5] R. Vagnetti, N. Camp, M. Story, K. Ait-Belaid, S. Mitra, S. Fowler Davis, H. Meese, M. Zecca, A. Di Nuovo, and D. Magistro, "Social robots and sensors for enhanced aging at home: Mixed methods study with a focus on mobility and socioeconomic factors," *JMIR aging*, vol. 7, p. e63092, 2024.

[6] J. Y. Choi, S. H. Lee, and S. Yu, "Exploring factors influencing caregiver burden: a systematic review of family caregivers of older adults with chronic illness in local communities," in *Healthcare*, vol. 12, no. 10. MDPI, 2024, p. 1002.

[7] P. Asgharian, A. M. Panchea, and F. Ferland, "A review on the use of mobile service robots in elderly care," *Robotics*, vol. 11, no. 6, p. 127, 2022.

[8] K. Zoughalian, J. Marchang, and A. Di Nuovo, "Access control architecture of assistive robots for physical activity wellbeing data," in *2024 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2024, pp. 1–7.

[9] S. N. Moutsis, K. A. Tsintotas, and A. Gasteratos, "Pipto: precise inertial-based pipeline for threshold-based fall detection using three-axis accelerometers," *Sensors*, vol. 23, no. 18, p. 7951, 2023.

[10] P. De Raeve, M. Azrour *et al.*, "Bluetooth-low-energy-based fall detection and warning system for elderly people in a smart home environment," *Journal of Sensors*, 2022.

[11] L. Pfyffer, R. Büchel *et al.*, "Development of an assistance robot for fall detection and reporting in healthcare," in *Proceedings of the International Conference on Robotics in Healthcare*, 2025, available at: https://www.researchgate.net/publication/391494832_Development_of_an_Assistance_Robot_for_Fall_Detection_and_Reporting_in_Healthcare.

[12] S. U. Ahamad, M. Ataei, V. Devabhaktuni, and V. Dhiman, "Omobot: a low-cost mobile robot for autonomous search and fall detection," in *2024 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2024, pp. 453–460.

[13] A. Charfi *et al.*, "Pose estimation-based fall detection using yolov8," *Sensors*, vol. 22, no. 15, p. 5599, 2022.

[14] G. Jocher and J. Qiu, "Ultralytics yolo11," 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[15] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," arXiv preprint, 2022, https://arxiv.org/abs/2212.04356.

[16] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.