Ref. Ares(2025)4364027 - 30/05/2025

D5.2 Processing infrastructure definition and setup for massive RF data

Grant Agreement Number	101099491
Action Acronym	HOLDEN
Action Title	Ethical Design of Holography with Dense wireless Networks (HOLDEN)
Funding Scheme	HORIZON-EIC-2022-PATHFINDEROPEN-01
Version date of the Annex I against which the assessment will be made	13/12/2022
Start date of the project	1/6/2023
Due date of the deliverable	31/05/2025
Actual date of submission	30/05/2025
Responsible	CNR
Contributors	CNR, AAL, ADANT
Dissemination level	Public

European Innovation Council

Funded by the European Union

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement No: 101099491.

Authors in alphabetical order

Full Name	Organisation	E-mail
Riccardo Bersan	ADANT	riccardo.bersan@adant.com
Sanaz Kianoush	CNR	sanaz.kianosuh@cnr.it
Nicola Pedrocchi	CNR	nicola.pedrocchi@cnr.it
Daniele Piazza	ADANT	daniele.piazza@adant.com
Song Yuqing	AALTO	yuqing.song@aalto.fi
Vahid Rouzegar	ADANT	vahid.rouzegar@adant.com
Stefano Savazzi	CNR	stefano.savazzi@cnr.it
Stephan Sigg	AALTO	stephan.sigg@aalto.fi
Cesare Tonola	CNR	cesare.tonola@cnr.it

Change History

Version	Date	Status	Author (Company)	Description
1.0	15/03/2025	Draft	CNR	First draft
2.0	01/04/2025	Revision	ALL	Revised version
3.0	06/05/2025	Revision	CNR	Revised version
4.0	12/05/2025	Final revision	ALL	Final version

Executive Summary

The deliverable highlights the components of the processing infrastructure, and the setups proposed for massive RF data collection. The application scenarios considered are twofold:

1) industrial internet of things and robotized environment (human-robot collaborative spaces)

2) healthcare, assisted living and smart spaces

The following contents are covered in this report:

- Short introduction of the HOLDEN project and the collaboration partners
- Description of the measurement setup and the data processing infrastructure in the selected CNR robotized environment
- Description of the measurement setup and RF data processing infrastructure in the Adant testhouse environment.

The architecture and infrastructure considered in this deliverable will be adopted in WP6 to validate the signal processing and hologram frame management tools.

Table of Contents

Abbreviations

Abbreviation	Description
3D	Three-dimensional
Aalto	Aalto University
BPA	back-projection algorithm
CNR	Consiglio Nazionale delle Ricerche
DOA	Direction of Arrival
EC	European Commission
EM	electromagnetic
ESM	Ethics Status Monitor
EU	European Union
HOLDEN	ethical design of holography in dense wireless networks
MUSIC	Multiple Signal Classification
PPDU	Physical Packet Data Unit
POLIMI	Politecnico di Milano
RF	radio frequency
TOI	target of interest
TUM	Technical University of Munich
TWE	University of Twente
UAV	uninhabited aerial vehicle
VNA	vector network analyzer
WP	work package
2D-FFT	two-dimensional Fast Fourier Transform

RDM	Range-Doppler Map
MTI	Moving Target Indication
CFAR	Constant False Alarm Rate
SoC	System on Chip
CW	Clockwise
CCW	Counterclockwise

1. Introduction

1.1. About HOLDEN

The ubiquitous perception by sensing of objects, subjects and gestures is a pivotal challenge for future technology: it enables personalized services such as smart living, automated logistics or interaction through free-space gestures. However, it also challenges ethical and moral boundaries and threatens privacy. HOLDEN proposes a radically new approach to perception by concisely analysing ethical constraints and privacy risks while re-thinking RF-based sensing. We establish necessary conditions for privacy preserving and ethically compliant sensing and develop new paradigms respecting these constraints.

For the first time ever, HOLDEN constitutes a concentrated effort to explore social aspects of RF-sensing to guide the technological advance and to derive technology for ethically and privacy compliant perception. Central to HOLDEN is the development of ethical and privacy constraints. We use these findings to derive privacy and ethically compliant concepts for RF-based perception. We will develop a system of distributed multi-antenna devices for simultaneous multitarget recognition and ubiquitous perception with unprecedented accuracy, which constitutes a radical paradigm shift from a technology-centric perspective to a privacy-centric one via privacy by design.

HOLDEN achieves this goal along three high risk, complementary, and privacy-centric paths:

Path 1: Continuous-space measurement points: Radio-based 3D vision by holographic image processing of RF wavefronts.

Path 2: Discrete-space measurement points: Advanced 3D beamforming for human-scale recognition and tracking through dense massive, connected antenna arrays.

Path 3: Signal processing and learning: High-dimensional tensor processing for the distinction of complex activities and motion from massive-dimensional RF data. The resulting breakthrough approaches and algorithms will be compared against application-level benchmarks via usage scenarios in the fields of logistics, smart living, and free-space

1.2. Partners

The consortium consists of four academic partners and a high-tech SME partner: (a) Aalto University (AALTO), Finland, (b) Technical University of Munich (TUM), Germany, (c) Consiglio Nazionale Ricerche (CNR), with third party Politecnico di Milano (POLIMI), Italy, (d) University of Twente (TWE), Netherlands, and (e) Adant (Adant), Italy. This consortium features the specialized and complementary expertise required to achieve the project objectives. AALTO will be responsible for the project management (WP1), covered by an experienced and dedicated project manager. Ethical aspects (WP2) will be addressed by TWE who is a pioneer in the field. Eventual gender differences in the ethical perception will be taken into account. TUM pioneered RF holography, which makes the ideal leader of WP3. In advanced distributed signal and information processing, CNR has more than 14 years of experience. CNR will lead WP4. Since more than 10 years, AALTO is active in radio sensing and machine learning based activity recognition. This expertise makes AALTO the ideal leader of WP5. Adant will contribute to the market analysis, application possibilities, and validation (WP6). Led by AALTO, dissemination with the website as one the media will be addressed by all partners. All academic partners are committed to early publication of results, e.g., via arXiv (open science).

2. Robotized environment setup and infrastructure

This section describes the robotic cell made available by CNR and used for data collection and testing of the gesture recognition algorithm conducted by Aalto (AAL).

2.1. Robotic infrastructure and application use case

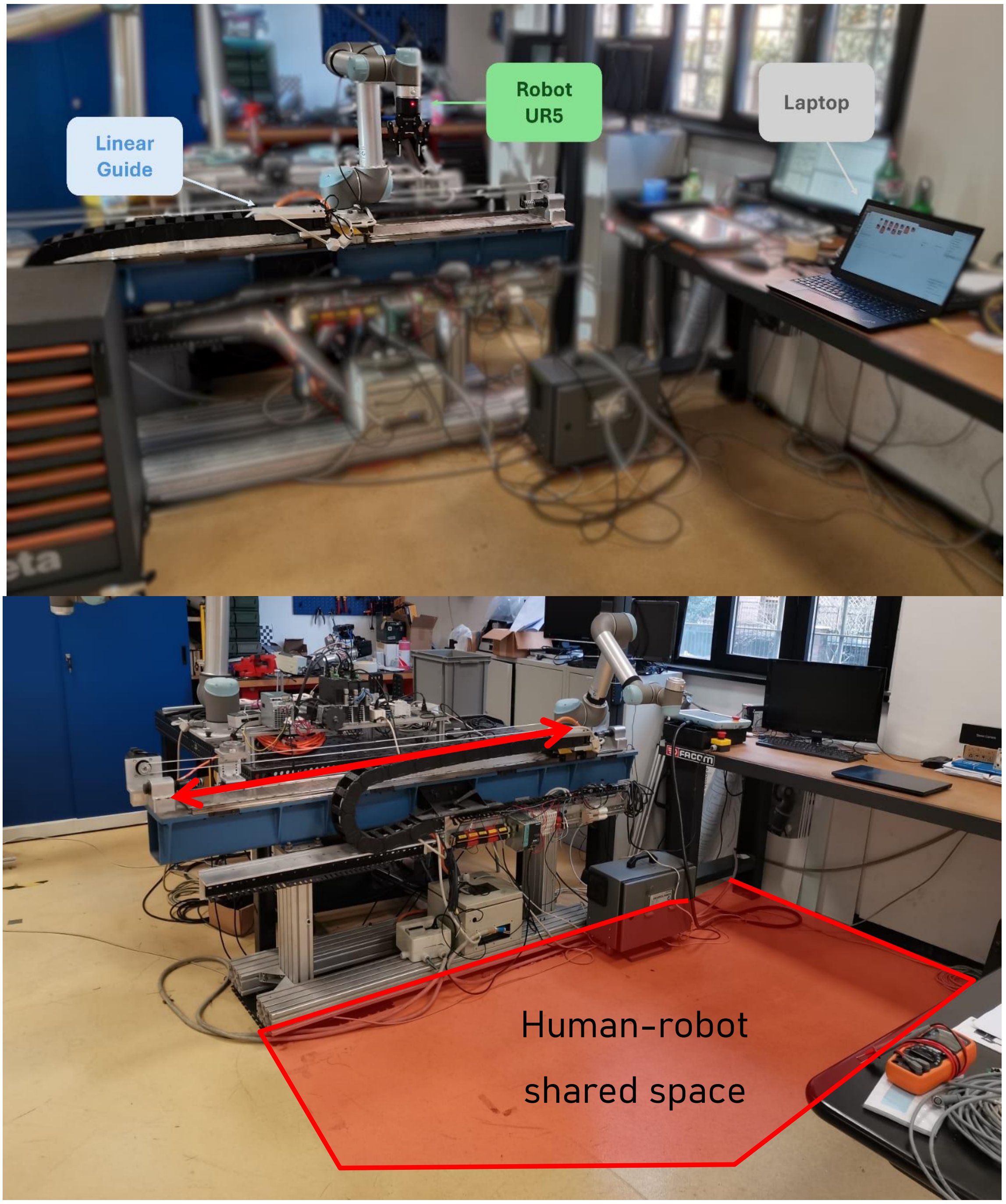

Fig. 1. Robotic environment (top) and highlighted human-robot shared space (bottom)

As robots become increasingly prevalent in both homes and industrial settings, the demand for intuitive and efficient human-machine interaction continues to rise. Gesture recognition offers an intuitive, hands-free control method and can be implemented using various sensing technologies—wireless solutions being particularly flexible and minimally invasive. While camera-based vision systems are commonly used, they often raise privacy concerns and can struggle in complex or poorly lit environments. In contrast, radar sensing preserves privacy, is robust to occlusions and lighting, and provides rich spatial data such as distance, relative velocity, and angle. These advantages make radar a strong candidate for gesture recognition and related tasks like path planning, mapping, and object tracking [3].

Motivated by these considerations, the robotic cell shown in Figure 1 was designed to simulate a realistic industrial scenario in which a human operator interacts with a robot using radar-based gesture commands. In this setup, the operator controls a typical pick-and-place operation—guiding the robot to pick up an object and move it to a new location—entirely through gestures.

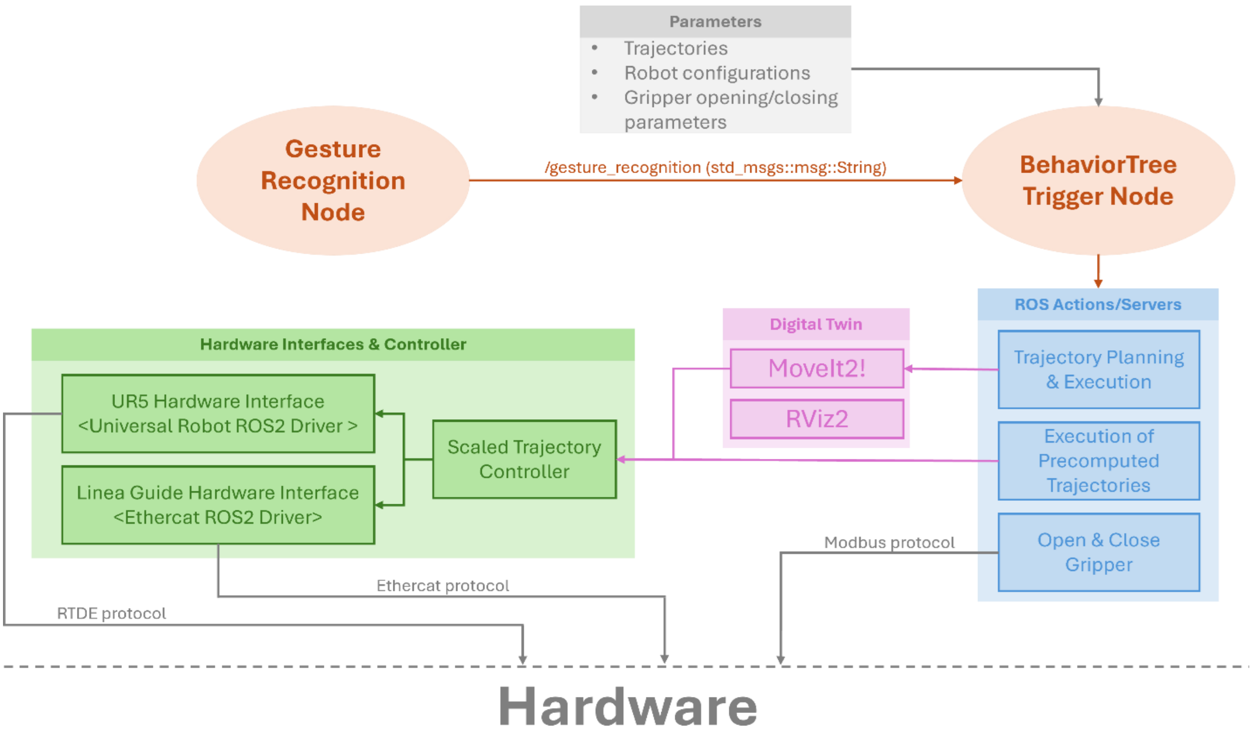

Fig. 2. Software architecture

Each gesture is linked to a specific robot action, so the operator must perform the correct sequence of movements to complete the task successfully.

The robotic cell consists of the following components:

- Universal Robots UR5 collaborative robot with a Robotiq 2f-85 gripper
- Linear Guide
- Computing Unit (laptop)

The UR5 is a widely used collaborative robot arm with 6 degrees of freedom. In the considered setup, it is mounted on a linear guide that adds a translational degree of freedom, extending both the robot's range of motion and its effective workspace. The entire robotic structure (UR5 + linear guide) is controlled via the computing unit (laptop) shown in Figure 1.

Figure 2 presents the software architecture used to control the robotic cell. The entire control stack is based on the following third-party software frameworks and libraries:

- **ROS 2 Humble [4]**: ROS 2 (Robot Operating System) is a widely adopted middleware consisting of software libraries and tools for building robot applications. It enables modular design by allowing software components (ROS nodes) to communicate through topics, services, and actions.
- **ros2_control [5]**: A ROS 2 framework for real-time robot control. It provides abstractions of the robotic hardware, known as hardware interfaces, which allow reading joint states (such as position, velocity, and acceleration) and sending control commands. The actual data exchange with the physical hardware is handled through hardware specific communication protocols. These interfaces connect to controllers that generate appropriate commands based on sensor data and predefined control logic.
- **BehaviorTree.CPP [6]**: A C++ library for building and executing Behavior Trees—an efficient and modular way to handle task switching in autonomous systems.
- **MoveIt 2 [7]**: A manipulation platform for ROS 2 that provides advanced tools for motion planning, manipulation, and control.

The architecture includes a dedicated ROS node, the *Gesture Recognition Node*, developed by Aalto University. This node processes radar signal data to recognize human gestures using a machine learning algorithm. The output is a string identifier representing the detected gesture, which is published on the topic "*/gesture_recognition*".

A second node, the *BehaviorTree Trigger Node*, subscribes to this topic. When a gesture message is received, the Behavior Tree corresponding to the string identifier is triggered. The execution of the Behavior Tree is managed using the BehaviorTree.CPP library [6]. Each Behavior Tree can define a single skill (i.e., a single action or motion) or a sequence of simple skills to accomplish a more complex task. These skills typically involve calling ROS services or actions that:

- Plan a trajectory to reach a desired robot configuration or pose
- Execute precomputed trajectories
- Open or close the gripper

Each service/action call is parameterized according to the specific task (e.g., the target configuration, the trajectory to execute, or how much to open/close the gripper). Trajectories are planned online using MoveIt 2 [7], which utilizes a digital twin of the robotic cell (Figure 3) to compute collision-free paths. Motion execution is handled by the *Scaled Trajectory Controller*, which performs fine-grained interpolation of the planned trajectory and sends real-time commands to the hardware interface. Additionally, the controller supports dynamic scaling of execution speed, allowing the robot's velocity to be adjusted from 0% (full stop) up to 100% (nominal trajectory speed).

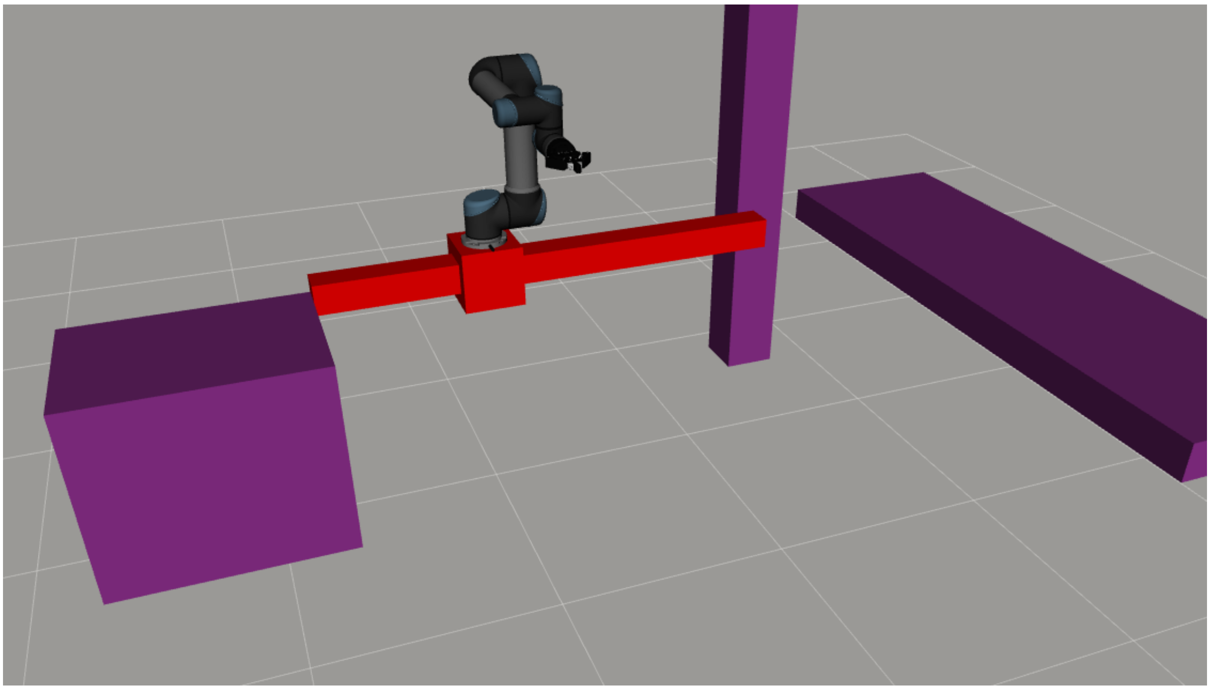

Fig. 3. Digital twin of the robotized infrastructure

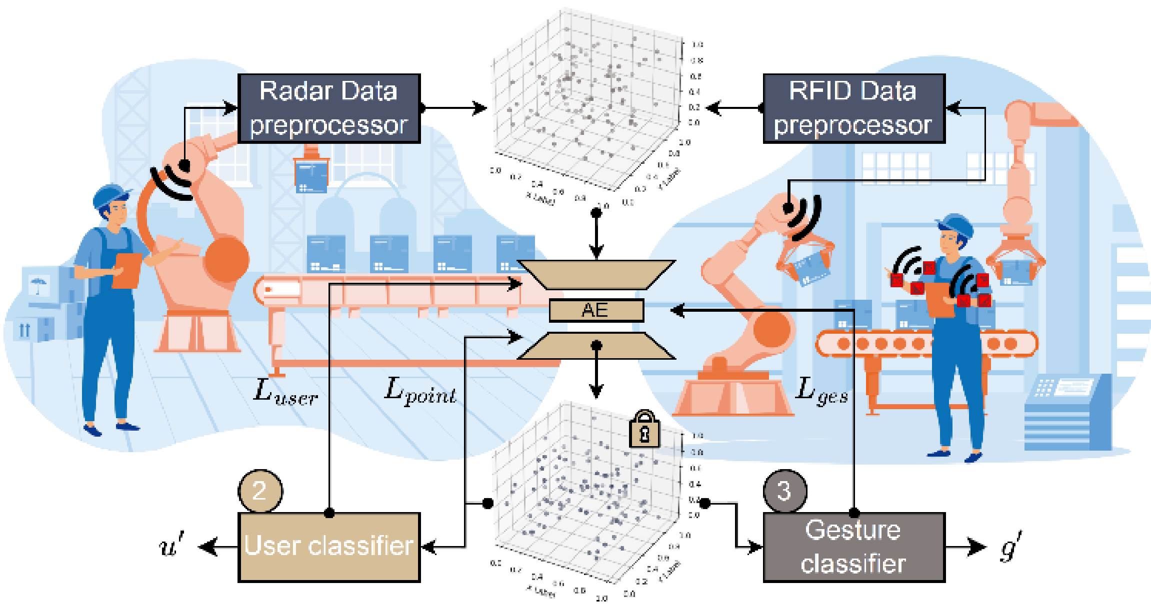

Fig. 4. Schematic view of the proposed unified privacy-preserving framework for RF sensing modalities that utilize point cloud representations. The framework is shown for radar and RFID tags.

Fig. 5. Gesture recognition testing in the CNR robotic lab.

2.2. Gesture recognition framework and processing infrastructure

The body gesture recognition system is based on radar-based point cloud data representations, which enable efficient spatial filtering options. As highlighted in D4.3, several radar designs have been successfully deployed and installed targeting precise half and full body gesture, pose and activity recognition.

A data processing pipeline is designed to obtain the point cloud data representation. The process begins with raw radar data in a 3D matrix of size $N_c \times N_{chirp} \times N_{chan}$ (number of channels × number of chirps × number of antenna elements). After reshaping, a 2D-FFT is performed on each of the four channels, and their outputs are summed to generate an RDM. MIR and CFAR detection are applied to the RDM to filter out static clutter and isolate moving gesture targets. For each detected target point, an FFT is performed across the N_{chan} dimension to obtain the angle spectrum, forming an angle map. Using the range and estimated angle, the target's Cartesian coordinates are computed. For each frame, the result is a list of detected objects with features including peak (signal strength), range (distance), Doppler (radial velocity), x, and y — ready for further analysis or classification. The complete processing chain is implemented on-chip and runs on the radar SoC.

To simplify the algorithm and accelerate the gesture recognition process, a one-dimensional convolutional neural network is applied. The maximum number of frames is 50, and the maximum number of detected objects per frame is 65, resulting in each gesture sample being represented as a 50 × 325 matrix. The input passes through a series of convolutional layers, ReLU activation functions, dropout layers, and max-pooling layers, and finally produces classification results. The entire process is implemented in real time using buffer management.

Table 1: Composition of the dataset

Settings	Samples
only hand in front of the radar	≈2000/class
hand and human in front of the radar	≈200/class
robotic arm at the back of the human	≈200/class

Table 2: Confusion matrix

True label \ Predicted label	NONE	UP	DOWN	LEFT	RIGHT	CW	CCW	Z	S	X
NONE	0.94	0.006	0.018	0.006	0.003	0.024	0.003	0	0	0
UP	0.013	0.97	0.0027	0	0.011	0.0027	0	0	0	0
DOWN	0.0054	0	0.99	0	0	0.0027	0	0	0	0
LEFT	0.016	0	0.0027	0.98	0	0.0053	0	0	0	0
RIGHT	0	0.0027	0.0027	0	0.99	0	0	0.0027	0	0
CW	0.0053	0	0	0	0	0.99	0	0	0	0
CCW	0.0081	0.0027	0.0027	0	0	0.0027	0.98	0.0027	0	0
Z	0.00081	0	0.0054	0	0	0	0.0027	0.97	0.016	0
S	0.013	0	0	0	0	0.0081	0.0027	0.0054	0.97	0
X	0.0027	0	0.0027	0	0	0	0	0.0027	0.0027	0.99

Initially, the dataset consisted of 9 gesture classes, with approximately 2,000 samples per class. However, all samples were collected under highly controlled conditions, where only the hand was present above the radar, with minimal background interference. Although the training performance appeared very promising, during field testing, significant interference from nearby humans and electronic devices was observed when the radar was deployed in a real-world environment. Therefore, additional data were collected under realistic conditions and incorporated into the dataset. The composition of the dataset is shown in Table 1. After retraining the model with the updated data, test results improved significantly, indicating that the gesture recognition system was now robust and ready for practical deployment. The confusion matrix is shown in Table 2. And the updated evaluation metrics are as follows: Accuracy: 0.9368, Recall: 0.8426, F1-score: 0.8544.

In the proposed case study example of Figure 5, the human guides the robot through a pick-and-place task by controlling the sequence of actions using gesture commands only. In particular, based on the hardware and processing infrastructure previously described, three different tests have been carried out:

- ❑ **Test 1:** The human performs specific gestures to activate each individual robot action required to complete a full pick-and-place task with a glass.
- ❑ **Test 2:** Same as test 1, but the robotic arm is behind the human, introducing interference.
- ❑ **Test 3:** The human uses gestures to trigger a sequence of robot actions that pour water from a bottle into a glass.
- ❑ **Test 4:** The human uses the "S" gesture to emergency stop the robotic arm.
- ❑ **Test 5:** The human continuously controls the positioning of the linear guide through hand movements (velocity control).

More specifically, in Tests 1 and 2, the gesture sequence was as follows:

- Performing the "swipe right" gesture moves the robotic arm to the right.
- "Swipe counterclockwise" opens the gripper.
- "Down" moves the arm downward.
- "Swipe clockwise" closes the gripper.
- "Swipe left" moves the arm to the left.
- The "S" gesture triggers the object placement operation.
- Finally, the "up" gesture commands the robotic arm to return to its home position.

In Test 3, a more complex task involving multiple objects was performed:

- The "swipe right" gesture moves the robotic arm to the right.
- The "down" gesture initiates the grasping of a glass.
- "Swipe left" moves the arm to the left.
- The "Z" gesture places the glass on the table.
- The arm then moves right again.
- The "X" gesture commands the arm to pick up a bottle.
- After that, it moves left once more.
- The "CW" gesture places the bottle next to the glass.
- The "CCW" gesture triggers the pouring action to pour water into the glass.
- Finally, the "up" gesture returns the robotic arm to the home position.

All tests were successfully completed, and a demonstration video was recorded for each scenario. And the interference in Test 2 did not influence at all.

In future work, we will further explore a privacy-preserving gesture recognition framework based on a graph autoencoder architecture that combines message-passing neural networks and multihead self-attention. The framework is shown in Figure 4. This end-to-end system is trained using adversarial techniques and multiple loss functions to retain task-relevant features while suppressing user-identifiable information. As part of ongoing development, we will validate the integrated unlearning mechanism and assess the framework's ability to maintain high gesture recognition accuracy while effectively mitigating identity-related privacy risks.

3. Test-house environment setup and infrastructure

3.1. Test-house environment

In this section we introduce the test-house infrastructure of ADANT partner and outline the planned tests for the measurement campaign. We thus introduce the planned test types, set up and configurations of the Channel State Information (CSI) processing infrastructure.

WiFi CSI data are collected in the testhouse environment and are used for testing algorithm implementations as well as optimized recognition beams. The goal is to recognize specific movements and locations and improve machine learning algorithms for data classification.

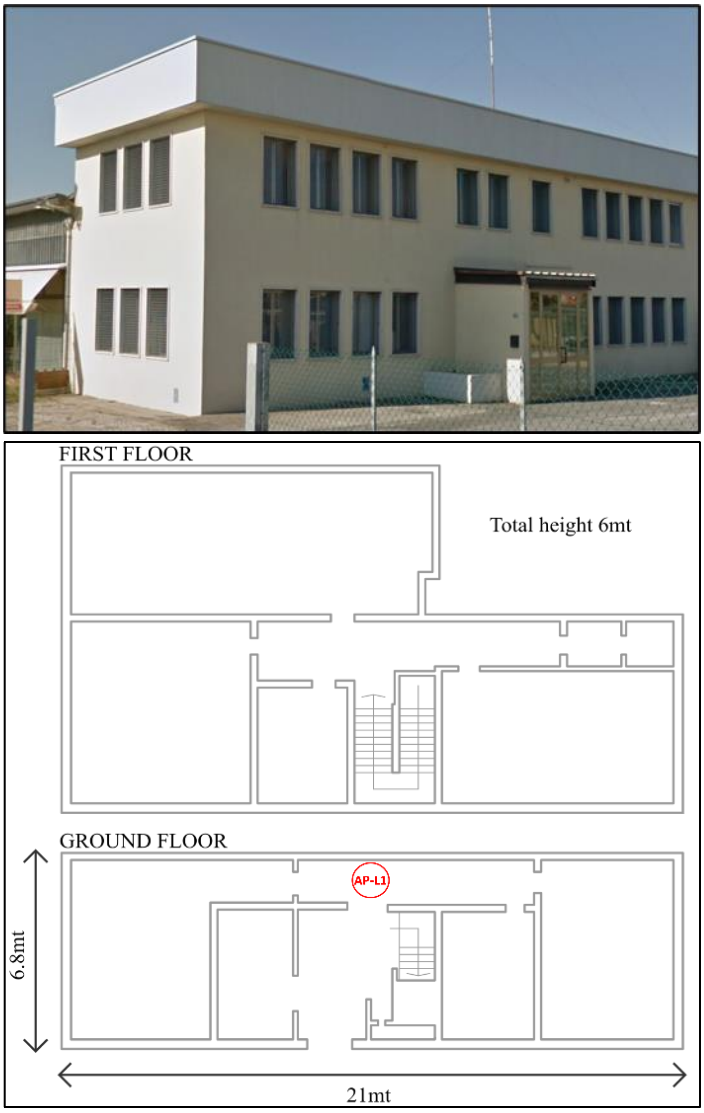

Fig. 6. Testhouse environment

The testhouse environment (see Figure 6) is located at Selvazzano Dentro – Padova and consists of two-story structure of approximately 340 sqft. In particular the following features are highlighted:

- ❑ **Pristine Wi-Fi spectrum free of interference**
- ❑ **State of art equipment for:**
  - OTA Wi-Fi and IoT fully automated tests
  - Wi-Fi standards (e.g., Wi-Fi 6, 6E, 7)
- ❑ **Configuration**: 2 floors, 11 rooms, 2 corridors, 2 stairs (see Figure 7)

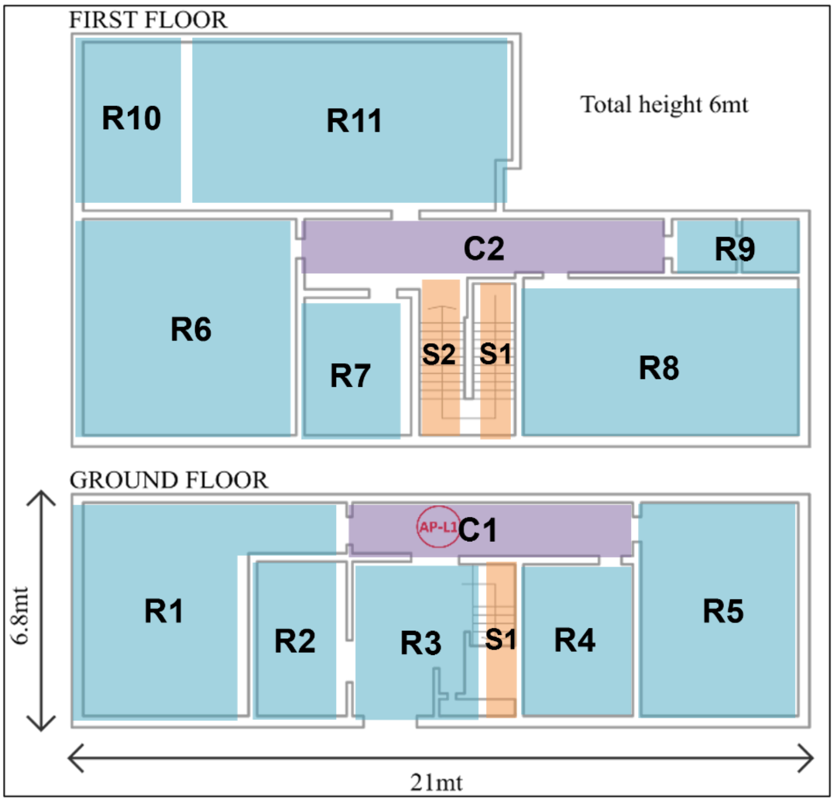

Fig. 7. Floor map of the testhouse environment

3.2. General test setup and processing infrastructure

The planned tests in WP6 will be performed in ADANT test house using Wi-Fi AP and Wi-Fi clients with the following characteristics:

- ❑ Number of Wi-Fi Stations (STAs) = 4
- ❑ Number of Access Points (APs) = 1

Wi-Fi clients can connect at both 5 and 6 GHz radios. The tests can be also performed at:

- ❑ 5 GHz: e.g. Channel 36, Bandwidth 160 MHz

- 6 GHz: e.g. Channel 33, Bandwidth 160 MHz

The AP extracts the CSI in the form of Channel State Information matrix. With respect to CSI data collection (see also D4.2):

- CSI collection duration for training will be 30 minutes for each test, with the frequency of 250 ms, corresponding to 4 samples/sec for each STA (~30K samples for each test)
- The AP is equipped with Smart Antenna which can change the antenna pattern (see D4.3) and allow testing different configurations. In particular, CSI collection will be performed with 2 up to 4 different antenna radiation patterns for each test (~120K samples per test)
- Data can be collected in CSV files or MAT files. In a second stage, they will be real-time processed via MQTT protocol

The selected main use cases are the following:

- Room localization: to detect and infer the presence/movement of individuals in and out of rooms using CSI analysis (time-frequency-beampattern analysis).
- People counting (counting the number of individuals and estimate the density and flow of people in the building): to obtain an "infrastructural snapshot" of the environment, namely an estimation of the number of subjects moving in the considered room over consecutive time instants
- Activity recognition (examples. wandering, walking, running, falling): to identify specific activities (e.g., walking, standing, or falling) using CSI variations.

For each use case the following performance metrics will be monitored by the proposed processing infrastructure, namely:

- Room localization:
    - Ability to find the room (localization) in which we have people standing or walking (accuracy/precision).
    - Ability to perform person re-identification, namely to classify an individual occupying the room based on macroscopic physical characteristics such as height, weight, and size or posture, walking pattern, gait or style (accuracy of person re-indentification).
- People counting
    - Conventional performance benchmarks: accuracy, precision, recall, AUROC
    - Statistical parity, equalized odds or predictive equality (for the definitions see D4.3) which measures possible biases in people counting based on the physical characteristic of the subject(s) being monitored
- Activity recognition:

- Conventional performance benchmarks: accuracy, precision, recall, AUROC
- Statistical parity, equalized odds or predictive equality (for the definitions see D4.3) which measures possible biases in activity detection based on the physical characteristic of the subject(s) being monitored
- Machine unlearning accuracy (see also D4.3)

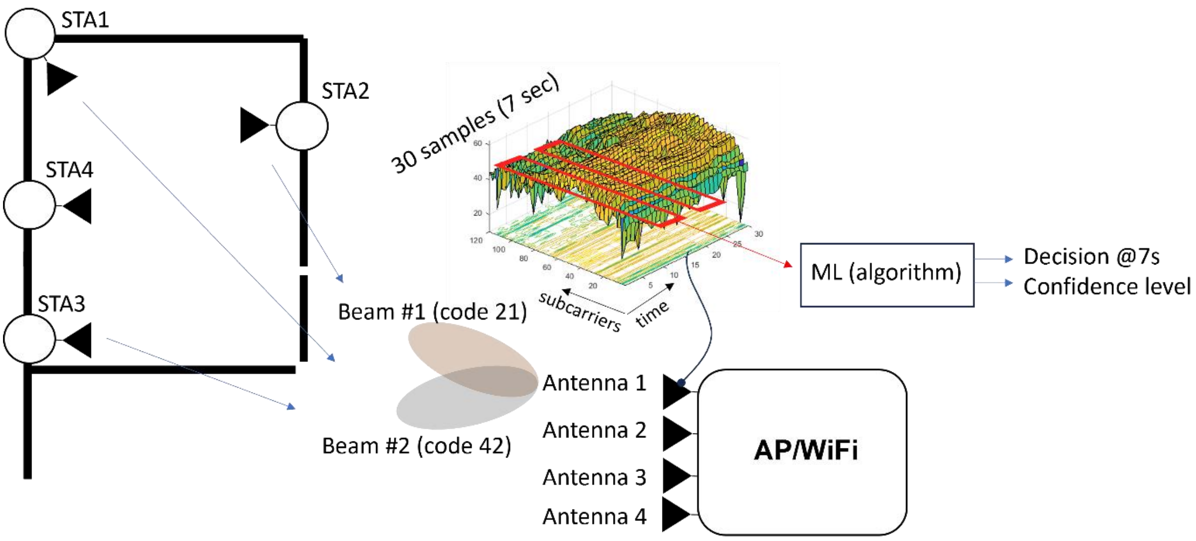

Fig. 8. Uplink CSI processing example (processing and learning on the AP)

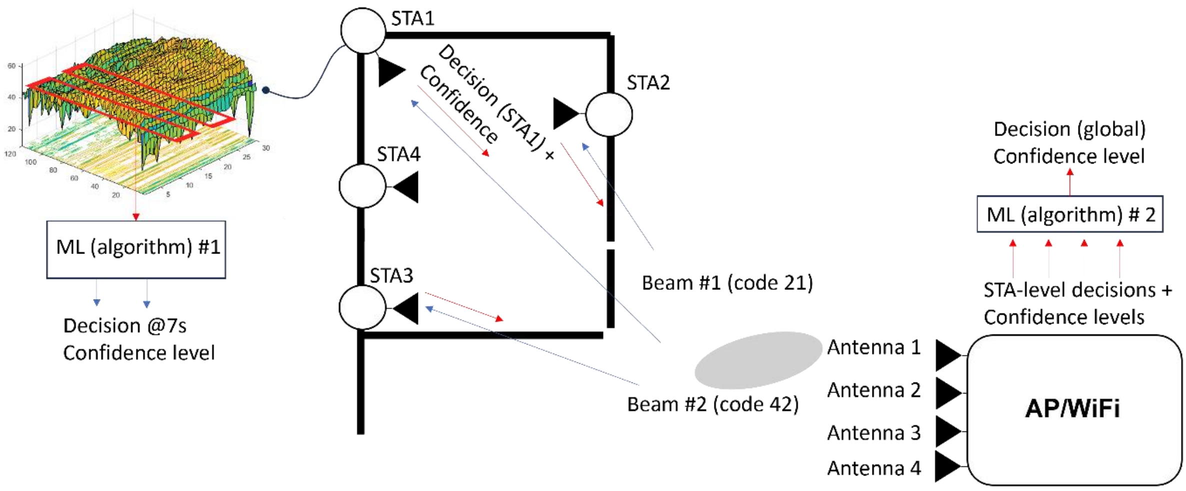

Fig. 9. Split-learning approach: processing and learning on the device and the AP

3.3. CSI processing architectures

Two data processing architectures are envisioned, described as follows

- **Data fusion on the AP**: The first one, shown in Figure 8, involves data collection at the Wi-Fi AP (over uplink) and subsequent processing through the training of a single AI model.
- **Split learning and inference**: The second one, depicted in Figure 9, proposes a split learning architecture where a portion of the AI model is trained locally on each Wi-Fi device. The results of this processing are collected by the Wi-Fi AP and then fed to a second stage. A second algorithm/model is this trained independently. The split learning architecture is attractive due its scalability potential and support to massive RF data processing.

For all cases the CSI information is collected over time for each Wi-Fi AP device, antenna at the AP (up to 4 antennas are available), beam steering profile (2 beampatterns per antenna). Information about the time-varying modulation and coding scheme (MCS) is also available for each PPDU frame. Further details are in D4.2.

4. References

[1] NanoRFE VNA6000-A. [Online]. Available: https://nanorfe.com/vna6000.html.

[2] S. Savazzi, V. Rampa, S. Kianoush and D. Piazza, "Pattern reconfigurable antennas for passive motion detection: WiFi test-bed and first studies," *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Istanbul, Turkey, 2019, pp. 1-6, doi: 10.1109/PIMRC.2019.8904262.

[3] T. Maiwald, J. Gabsteiger, R. Weigel and F. Lurz, "Gesture Recognition to Control a Moving Robot With FMCW Radar," 2024 IEEE Radio and Wireless Symposium (RWS), San Antonio, TX, USA, 2024, pp. 105-108, doi: 10.1109/RWS56914.2024.10438564

[4] Macenski, Steven, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. 2022. "Robot Operating System 2: Design, Architecture, and Uses in the Wild." Science Robotics 7 (66): eabm6074. https://doi.org/10.1126/scirobotics.abm6074.

[5] *ros2_control* documentation. [Online]. Available: https://control.ros.org/humble/index.html.

[6] BehaviorTree.CPP. [Online]. Available: https://www.behaviortree.dev

[7] S. Ioan A. and C. Sachin, "MoveIt." [Online]. Available: https: //moveit.ros.org

5. Table of Figures