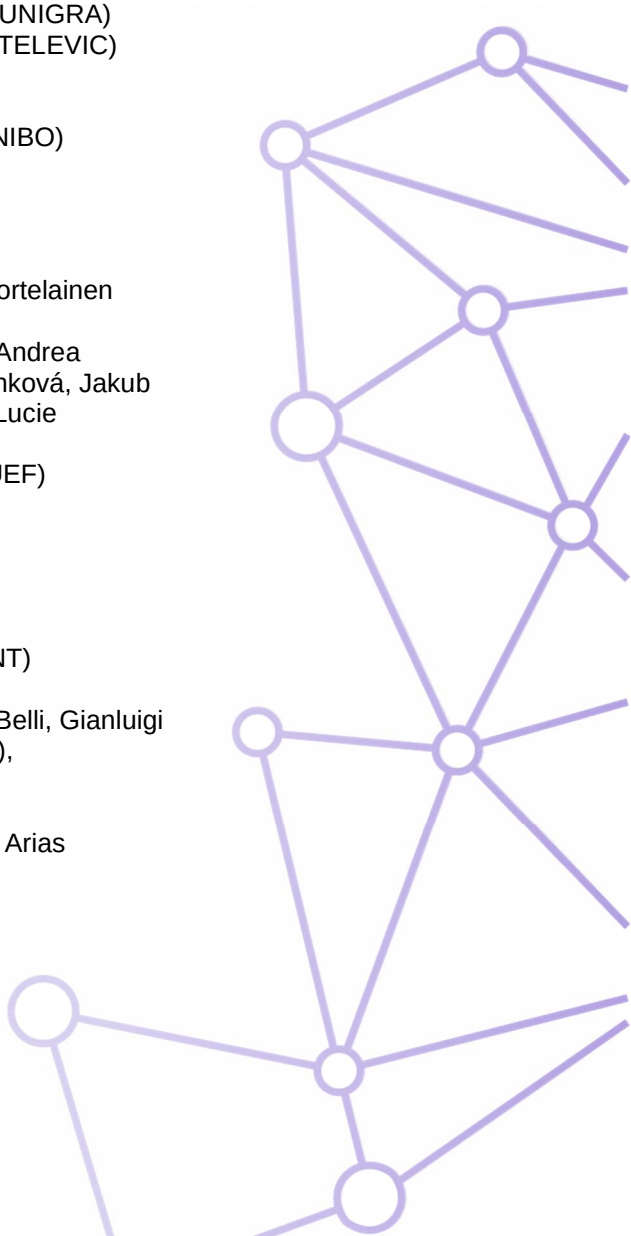
DistriMuSe

D5.3 Results of the first evaluation and stage 2 feedback

Lead Author: Raúl Santos de la Cámara (HIB)

Main editors:	Raúl Santos de la Cámara (HIB), Elena Muelas Cano (HIB), Roberta Presta (UNISOB)
Contributors:	Federica Viero, Alessandro Monteleone (RELAB) Alexandra Krassikova (IRBLL) Mikko Saajanlehto (ELIVE) Jesús Garrido, José Manuel Herrera (UNIGRA) Pieter Crombez, Dries Van Leemput (TELEVIC) Heico Sandee (SR) Vlastimil Benes (IMA) Roberto Girau, Lorenzo Bacchiani (UNIBO) Ljubomir Jovanov (imec) Kari Bäckman (BENETE) Moritz Weigand (TRILITEC) Abhijit Pal (UBREMEN) Maria Jokela,Titta Kemppi, Juha M. Kortelainen (VTT) Jaromír Hubálek, Daniela Chlíbková, Andrea Němcová, Luláš SMital, Helena Šimůnková, Jakub Arm, Martin Rosa, Radovan Smíšek, Lucie Šaclová, Tobiáš Goldschmidt (BUT) Samu Kainulainen, Teemu Laitinen (UEF) Andrea Generosi (EMOJ) Ludo Cuypers, Jade Guo (COMmeto) Stefan Schulte (FLIR) Raya-Zoe Nikol (CIT) Reda El Hail (KUL) Anton Lambrecht, Jorg Wieme (UGENT) Fokke van Meulen (KMPHG) Veronica Mattioli, Luca Davoli, Laura Belli, Gianluigi Ferrari (UNIPR), Saad Saleh (HOLST), Fernando Seco (CSIC) Geert Vanstraelen (Macq) Miriam Zambudio Martínez, Alejandro Arias Jiménez (ODINS)
Reviewers:	Enrico Ferrari (RULEX) Fernando Seco (CSIC) Michael Roelleke (Bosch)



Chips JU

EUROPEAN PARTNERSHIP

DistriMuSe

Version: 1.01 Release Version

Date: 2025-10-09

Due date of Deliverable: 2025-09-30

Actual submission date: 2025-10-09

Dissemination level: PU

Revision history				
Version	Date	Modifications	Authors	Status
0.1	02/07/2025	Initial ToC sent for comments	Raúl Santos (HIB), Elena Muelas (HIB), Roberta Presta (UNISOB), Juha.M.Kortelainen (VTT), Mikko Saajanlehto (ELIVE)	Draft ToC
0.5	16/09/2025	Intermediate contributions to sections 3, 4 and 5	All, Raúl Santos de la Cámara (HI Iberia)	Draft
0.9	24/09/2025	All contributions compiled. Creation of a review candidate document.	All	Review candidate
0.99	07/10/2025	Creation of final draft for last minute review, integration of all fixes post internal review.	All, Raúl Santos de la Cámara (HI Iberia)	Final draft
1.01	09/10/2025	Final version of the document.	Raúl Santos de la Cámara (HIB)	Release document

Table of content

List of Acronyms

Abbreviation or acronym	Meaning
AAL	Active and Assisted Living
ADAS	Advanced Driver Assistance Systems
ADASIS	Advanced Driver Assistance Systems Interface Specifications
AGV	Automated Guided Vehicle
AMR	Autonomous Mobile Robot
ANT+	
AROM	Active Range of Motion , as applied
B2B	Business to Business
BLE	Bluetooth Low Energy
C-ITS	Cooperative Intelligent Transportation System
CNN	Convolutional Neural Network
COTS	Commercial off-the-Shelf
D	Deliverable
Demo	Demonstrator
DrDT	Driver Digital Twin
DMS	Driver Monitoring Systems
DOS	Denial Of Service
DPIA	Data Protection Impact Assessment
DSS	Decision Support System
ECG	Electroencephalogram
FMCW	Frequency-Modulated Continuous-Wave
GDPR	General Data Protection Regulation
GIS	Gait Instability Score
GNSS	Global Navigation Satellite System
GS	Gait Speed
GSM	Global System for Mobile communications
GP	General Practitioner
GPS	Global Positioning System
GW	Gateway
HADM	High Accuracy Distance Measurement
HDOP	Horizontal Dilution Of Precision
HMD	Head-Mounted Device
HMI	Human-Machine Interface
HRI	Human-Robot Interaction
HR	heart rate
HRV	heart rate variability
I2I	Infrastructure to Infrastructure
IDS	Intrusion Detection System
IMU	Inertial Measurement Unit
IoT	Internet of Things
IoMT	medical Internet of Things devices
IR	Infrared
ITS	Intelligent Transportation System

KPI	Key Performance Indicators
LDA	Linear Discriminant Analysis
LIDAR	Light Detection and Ranging
M	Month
MCI	Mild cognitive impairment
ND	non-distracted
ML	Machine Learning
MOTA	Multiple Object Tracking Accuracy
MQTT	Message Queuing Telemetry Transport
NIR	Near Infrared
NASA-TLX	NASA Task Load Index
OBU	On Board Unit
OEM	Original Equipment Manufacturer
PAT	heart Pulse Arrival Time
PCA	Principal Component Analysis
PD	Parkinson's disease
PL	Power Line
POG	Point of gaze
PPG	Photoplethysmography
PRO	Patient-reported outcome
PROM	Patient-reported outcomes measures
PSG	PolySomnoGram as a medical sleep study
RDT	Road Digital Twin
RF	Radio Frequency
RGB	Red-Green-Blue
RISC	Reduced Instruction Set Computer
RSE	Road Side Equipment
RSU	Road Side Unit
RTMP	Real-Time Messaging Protocol
SAGAT	Situation Awareness Global Assessment Technique
SAM	Self-Assessment Manikin
SART	Situation Awareness Rating Technique
T	Task
TUG	Time Up and Go
VRU	Vulnerable Road User
VDT	Vehicle Digital Twin
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything
UC	Use Case
UEQ-S	User Experience Questionnaire
UI	User Interface
USP	User Services Platform
UWB	Ultra-Wideband
WBAN	Wireless Body-Area Network

WP	Work Package
XAI	Explainable artificial intelligence
YOLO	You Only Look Once

1 Preface

This document is a public deliverable of the DistriMuSe project. It provides a thorough overview of the validation of the technologies developed during the first cycle in a variety of health, automotive and robot cooperation use cases. Partners collaborated to integrate and validate the technologies and focused on assessing the feasibility of the technologies as corresponding to the first iteration, where the second iteration approach will be wider and incorporate an in-depth study of the user aspects.

As this is a public deliverable, some content was paraphrased from previous deliverables to provide more clarity and context to the presented validation outcome – the main purpose of D5.3. This repetition is presumed acceptable as it makes the document more self-contained.

The project continues with a slightly longer second cycle, which starts with a revisitation of the planning documents in WP1 which is to be influenced by the findings in this first cycle. This document plays a major role in presenting these lessons learnt from the first cycle, which will be used to adjust the focus in the second cycle. For that purpose, indications from all of the pilots upon the direction and required changes in the approach for the second cycle is presented. The validation in the second cycle will eventually lead to the follow-up of this deliverable, D5.6, which will report on the final pilots in the project. The last pilots will make more emphasis on user aspects such as acceptability and perceived impact on the user and will take place in more real-world settings where possible.

2 Introduction

This deliverable is the joint result of tasks T5.3 and T5.4 of WP5. In the following two subsections we present the separate objectives of each task, which have informed the structure of the document and the approach to be followed in the evaluations.

2.1 Task 5.3 objectives

Task T5.3 deals with the general evaluation of the demonstrators and pilots of the project. The text in the definition of the task reads as follows: "***validation of the integrated technologies*** *against the* ***overall project and technology KPIs****. Pilots are* ***realistic evaluations of integrated setups*** *that combine different demonstrators, executed in a controlled environment close to real-life applications. Evaluations will be performed in two rounds: the* ***first one focusing on technical feasibility****, and the second one addressing qualitative aspects*".

Thus, the objectives for this task are as follows:

- Explicitly present the target set-ups for evaluation (technical demonstrators and the real-world scenarios in which they are tested).
- Present the results of testing according to internal metrics but also to those corresponding to the project objectives and KPIs.
- Demonstrate the soundness and feasibility of the technical approaches chosen and lay the groundwork for more thorough testing including user aspects in cycle 2.

For the current understanding of this document, it is also important to distinguish the different terms used to name the project testable outcomes. As per the project proposal, these are as follows:

- **Domain** - The application domain where the envisioned solution will be used. The project distinguishes three domains:
    - *Health & Wellbeing* - In the health sub-domain we focus on monitoring of health conditions in hospital, at home, in care facilities or on the move.
    - *Mobility* - This domain focuses on automotive and mobility cases where people are at risk to be injured due to dangers in the traffic environment, their own behavioural patterns and psycho-physiological conditions.
    - *Robotics* – This is a subdomain of industrial environments, focusing on the use of robots in an environment where people are present
- **Use case (UC)** - A use case can be described by a scenario showing the activities of actors (people, machines) facing the challenges of a domain and illustrating how the envisioned solution is used in order to solve these challenges. It provides a context to the developments in the project and drives the requirements by real-world challenges.
- **Proof-of-Concept (PoC)** - A technical implementation of a sensor, platform, or algorithm to show its feasibility and test its features. It is not necessarily complete, neither can it be commercialised as such. These are the components resulting from the technical work packages.
- **Demonstrator** - A collection of integrated PoCs that partially or fully implements the concept described in a use case. The demonstrator will allow for the evaluation and/or validation of some aspects of the use case concept in a pilot. A set of demonstrator instantiations with similar purposes but differing in technical details (e.g. to cater for pilot specifics) can be referred to as **Demo**s.
- **Pilot** - An evaluation of a demonstration concept undertaken along with experts or prospective users. The pilots in this project will mostly aim at showing technical feasibility, obtaining information of their efficiency and accuracy, and gaining understanding of user acceptance with the additional goal of studying the user aspects of said technologies (e.g., acceptability, social impact and trust aspect).

In this document, the main divisions (top level chapters) distinguish between the three domains. Within them, we usually divide by Pilots, which is the most usual way of working in the DistriMuSe project, where pilot leaders have ample independence and organize work in their area with freedom. However, this division is sometimes imperfect and not exactly tailored to the needs of each domain. As such, in the Mobility oriented chapter 4, a more demonstrator-oriented division is favoured.

This divergence was detected late in the production process for this document to fully homogenise it. For the cycle 2 and subsequent deliverables, an attempt will be made to proposing a more homogenic division.

2.2 Task 5.4 objectives

Task 5.4 investigates user aspects across all use cases, demonstrators, and pilots. Building on the foundations established in D5.1, the task has now moved from planning and early design reflections to the first round of evaluation activities. Its main objective remains to ensure that design solutions are usable, acceptable, and ethically responsible for diverse end-users and stakeholders. Task 5.4,

running from M10 (February 2025) to M36 (April 2027), thus progresses in D5.3 from planning to implementation, presenting the first evidence of user validation within pilots. The reflections reported here will guide the refinement of pilot strategies and provide crucial input for the subsequent stages of evaluation, leading towards the synthesis of human-centered results and ethical insights in the final D5.6.

In particular, Deliverable D5.3 focuses on reporting the results of the first evaluation cycle (Cycle 1) and on integrating feedback for the planning of Cycle 2. Depending on the stage of development of each pilot, contributions either include first evaluation results or, where evaluations have not yet been carried out, the detailed plans for Cycle 2. This variability reflects the different maturity levels of pilot prototypes, the sequencing of technical integration activities, and the methodological choices best suited to each use case.

The work developed within this task is articulated across three complementary dimensions:

User aspects: stakeholder engagement in pilot development. This section reports whether and how pilots have involved stakeholders and target users during the development phase. Stakeholders are not limited to external actors but should extend beyond the core technical development team, while target users refer to the intended end-users of each solution. Contributions include engagement activities such as consultations, focus groups, observations, or testing sessions, aimed at collecting expectations, validating scenarios, or identifying aspects considered important by users and stakeholders. For each pilot, a concise factual description is provided, outlining who was involved, the type of engagement conducted or planned, and how the feedback has been considered in pilot refinement.

User-based KPI assessment. This section presents the evaluation of user-related KPIs defined in earlier phases (cf. D5.1). Pilots are invited to report on the assessment activities carried out during Cycle 1, including the methods used and any preliminary results related to key user aspects such as usability, user experience, situational awareness, and others. Where evaluations have not yet been conducted, pilots indicate whether they are planned for Cycle 2 and recall the type of assessment that will be performed. This ensures continuity with the KPI framework already established and provides a baseline for Stage 2 adjustments.

User aspects: gender/age issues and ethical concerns. This section documents the integration of ethical reflections into pilot development, in line with the Ethics Exercise Tool and WP7 guidance. Pilots state whether and when an ethics focus group has been conducted or is scheduled, who was involved in the discussion and the rationale for their inclusion. Detailed results of these discussions are reported in the dedicated WP7 deliverables, so this section is limited to factual reporting of the activity status. Where relevant, pilots may also indicate if initial insights from the ethics reflection have already informed their evaluation strategy or highlighted aspects to be addressed in Cycle 2.

2.3 DistriMuSe Use Cases, Demonstrators and Pilots: summary

What follows is a summary table of the different demos and their associated pilots proposed for DistriMuSe cycle 1, grouped into the different Use Cases (health monitoring, mobility and robot interaction monitoring). Each of the pilots works in an independent fashion towards fulfilling a subset of each Use Case objectives, proposing experiments in a given location. In the following subsections of chapter 3 -5 we will use this structure to further detail the activities for each part of the project.

Table 1 DistriMuSe UCs and related demos and pilots

Use Case	Demo	Pilot	Objective	Owner
UC1	Demo 1.1 - Human life-style monitoring	P1-LL	Monitor daily activities of early-stage MCI patients and elderly residents in care facilities.	HIB, IRBLL
		P1-KUO		VTT
		P1-G		TELEVIC
	Demo 1.2 - Sleep monitoring	P1-KMPHG	Develop less obtrusive methods for monitoring sleep and assessing sleep disorders by measuring vital signs in a novel way that could replace traditional and more obtrusive methods.	KPNHGE, TUE
		P1-KSL		UEF
	Demo 1.3 - Sports performance and health assessment	P1-BRNO	Monitor physical activity to extract various parameters and their relations with health, create warnings about potential health issues, and estimate performance levels and maximum effort using wearable sensors and ML-based processing.	BUT
		P1-TOR		UNITO
UC2	Demo 2.1.1 - Collaborative situational awareness simulation	P2-BRU	Enhance situational awareness collaboration in a traffic simulator by communicating digital twins, which integrate information from simulated vehicles and roadside units to create a combined traffic situation view.	MACQ
	Demo 2.1.2 - Vehicle situational awareness digital twin	P2-TMP	Integrate available information from on-vehicle sensors into a vehicle situational awareness digital twin that can share data with other traffic users and receive additional observations from them.	VTT
	Demo 2.1.3 - Vehicle and roadside unit collaboration	P2-HH	Illustrate sharing information on observed traffic situations using V2I communication between a real car and a roadside unit, involving other vehicles, VRUs, and road/infrastructure status data.	CIT
	Demo 2.2.1 - Driver distraction monitoring in a driving simulator	P2-REGG	Enhance road safety by using advanced digital twin technology to monitor and predict driver behaviour, utilising a driving simulator.	RELAB
	Demo 2.2.2 - Driver and passenger mood and distraction monitoring	P2-TMP	Monitor driver and passenger behaviour in a real vehicle and in a real traffic environment.	VTT
	Demo 2.2.3 - Driver attention and traffic situation matching	P2-TMP	Match internal and external vehicle situations by using on-vehicle sensors to monitor surroundings and track other road users, especially VRUs.	VTT
UC3	Demo 3.1 - Sensor fusion as a reliable safety measure	P3-BEST	Demonstrate reliable safety in human-robot interactions in industrial settings using sensor fusion and ML-based data analytics, with adaptive HMI strategies to enhance collaboration, optimise task performance, improve safety awareness, and reduce mental stress.	SR
	Demo 3.2 - Enhancing	P3-GRA	Use virtual reality simulations to identify risks, test human-machine interfaces, train AI, and optimise sensor placement in the design and operation of a	UNIGRA

	safety with virtual reality		robotic system, running in parallel as a digital twin to provide real-time risk alarms and additional imaging.	
	Demo 3.3 - Dynamic factory robots-human safe interaction	P3-SON	Develop and test a dynamic factory where mobile robots and humans work safely together, using a fleet management system and multi-sensor fusion to handle diagnostics, safety monitoring, job scheduling, and adaptability to changing environments.	PRODRIVE

3 UC 1 demonstrator evaluation cycle summary

Use Case 1 (UC1) is focused on improving continuous hybrid human health monitoring to provide users with personalized health metrics that can be used for health assessment and evaluation, as well as the prevention of emergencies. In particular, across three demonstrators and seven pilots, UC1 is focused on three aspects of human activities: daily activities, sleep, and sport performance.

Demo 1.1: Human lifestyle monitoring

Demo 1.1 focuses on monitoring daily activities of older adults using a multi-modal sensing and decision-making system consisting of a combination of visual, wearable, radar, and other sensors.

- P1-LL: Older adults with cognitive and mobility impairments and Parkinson's disease
- P1-KUO: Older adults living in their own apartments
- P1-G: Older adults living in care and assisted living facilities

Demo 1.2: Sleep monitoring

Demo 1.2 focuses on developing less obstructive and invasive sleep monitoring technologies that provide reliable results.

- P1-KMPGH: Patients diagnosed with sleep disorders
- P1-KSL: Healthy volunteers without diagnosed or self-reported sleep disorders

Demo 1.3: Sport performance and health assessment

Demo 1.3 focuses on wearable sensors monitoring physical activity to provide personalized insights into individuals' physical performance and overall health.

- P1-BRNO: Semi-professional athletes
- P1-TOR: Older adults, including individuals with neurological conditions

3.1 P1-LL evaluation cycle 1

3.1.1 Final pilot set-up

Pilot P1-LL is coordinated by HIB and IRBLL in collaboration with ACORDE, CSIC, EVALAN, and UVIGO. The aim of the pilot is to validate a multi-sensor and decision-making model for monitoring of daily activities of older adults. The pilot primarily targets older adults with neurodegenerative diseases, specifically mild cognitive impairment and Parkinson's disease, as well as their informal caregivers and healthcare providers. The ultimate goal is to provide patients and their care partners with personalized insights into older adults' health to identify emergencies and progression of health conditions. The pilot integrates a multi-sensor system consisting of RGB cameras for bradykinesia detection, microphones for voice and linguistic analysis, Internal Motion Units (IMUs) and RGB/depth

cameras for gait and posture analysis, and FMCW radar for activity and fall detection. The inputs are processed in real-time and combined into a decision meta-model providing users with interpretable health metrics. The study is conducted at the Biomedical Research Institute of Lleida in a Living Lab, a laboratory space designed to imitate a two-bedroom apartment with functional plumbing and wiring. The multi-sensor system within the Living Lab is illustrated in Figure 1.

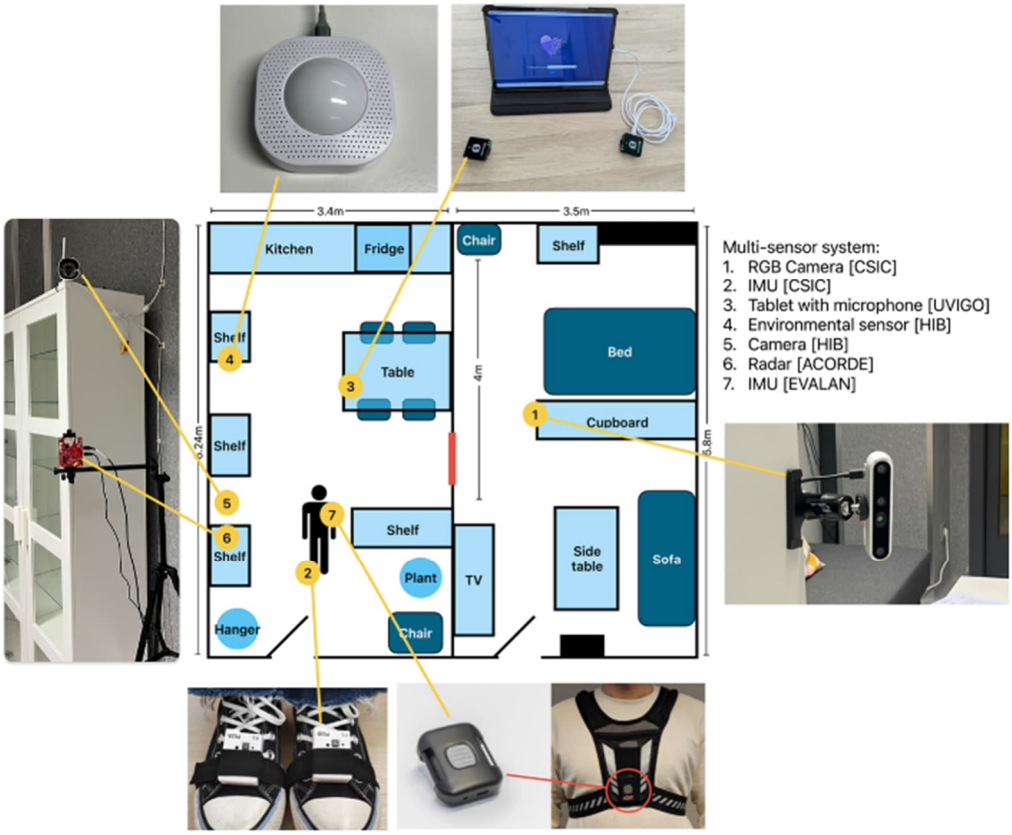


Figure 1 P1-LL Living lab and the sensors comprising the system.

The pilot is divided into two cycles. In cycle 1, the feasibility of all components of the multi-sensor system will be individually examined, following their respective validation and KPI plans, with the same study participants. Usability will be preliminarily explored through interviews with study participants. In cycle 2, the multi-sensor system will be validated as a whole, focusing on technological readiness, and end-user usability and acceptance will be examined more closely. The data and user feedback obtained in cycle 1 will be integrated in cycle 2 to improve the multi-sensor system.

Since the initial proposal in D5.1 and D5.2, several changes have been made. The target recruitment sample for cycle 1 was reduced to 15 individuals, split between three groups: older adults with mild cognitive impairment (n=5), older adults with Parkinson's disease (n=5), and healthy controls (n=5). The target recruitment sample for cycle 2 will remain 60 individuals. Moreover, ODINS joined P1-LL with the main aim of implementing security and authentication solutions, specifically a lightweight blockchain-based authentication mechanism for medical IoT devices or IoMT.

3.1.2 Pilot evaluation execution and protocol details

[1]To validate the multi-sensor system, three groups of individuals will be recruited for a series of tasks at the lab: 1) older adults with MCI (Mild Cognitive Impairment), 2) older adults with PD (Parkinson's Disease), and 3) older adults with no diagnosis of neurodegenerative diseases. The third group will serve as a control. Individuals will be asked to complete a series of tasks imitating activities routinely performed by older adults at home. These tasks are designed to evaluate individuals' functioning in basic and instrumental activities of daily living and are based on activities previously validated in other studies focused on detecting MCI and PD in older adults.

Participants will be instructed to perform all tasks in the outlined order. Instructions for each activity will be given to the participant by the researcher before each activity. Once the activity starts, no further instructions will be provided unless the participant explicitly asks for assistance. If asked for assistance, the researcher will give incremental support. In addition, as the last activity, participants will be asked to complete the Timed Up and Go (TUG) test. The TUG test is often used to assess mobility, balance, walking ability, and fall risk in older adults. In addition, the TUG can reliably detect mobility changes in individuals with Parkinson's disease.

The overall pilot architecture is pictured in Figure 1.

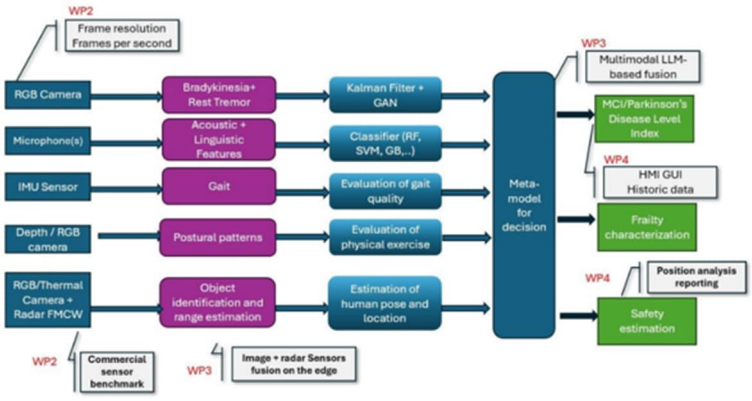


Figure 2 P1-LL Pilot architecture

The individual technologies that are integrated and tested in P1-LL are described below:

Bradykinesia and resting tremors detection [HIB]

We installed RGB cameras in the living quarters of the lab focused at the participants (patients and members of the control group). The cameras are placed in the living room and connected to an edge processing element (Raspberry Pi 5 B with the AI Hat including a Hailo Neural Network Accelerator). This is depicted in the Figure 3.

[1] The two paragraphs below are entirely derived from D5.2 (Chapter 2.1 P1-LL Implementation [HIB, IRBLL] / 2.1.1 Demonstrator 1.1) and describe the protocol details implemented in cycle 1.

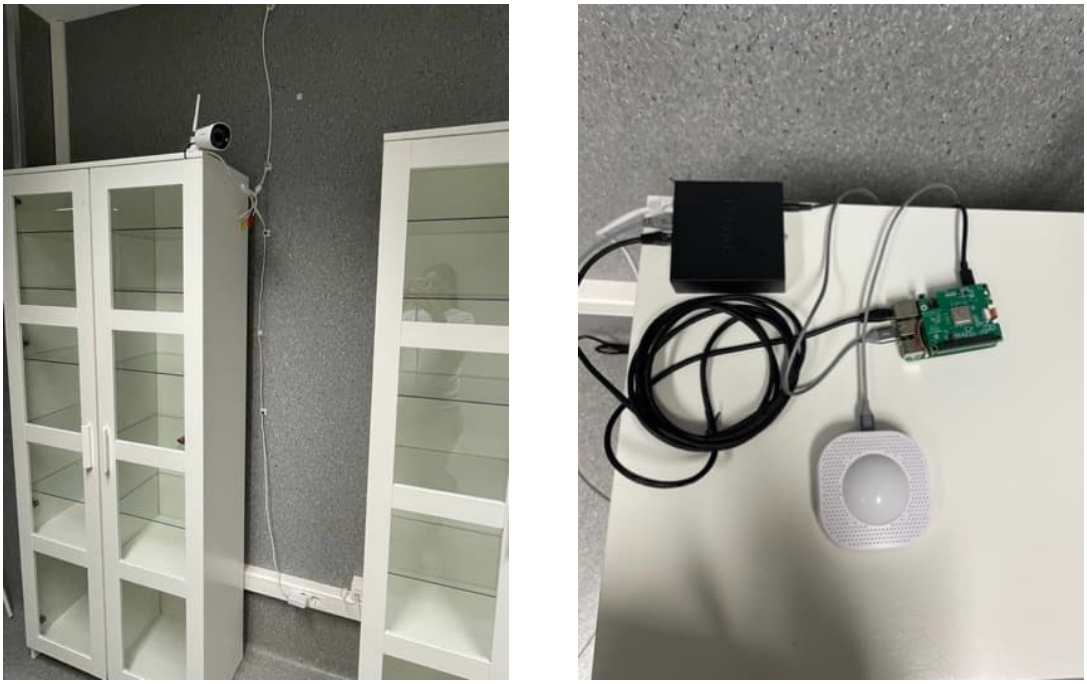


Figure 3 RGB camera (left) and edge processing element (right) for bradykinesia monitoring at IRBLL

The deployed system captures RGB picture frames that are processed in the edge element using computer vision algorithms. In addition, this is used to detect the presence of subjects in the living room. The results of the detection are sent using a Zenoh app that connects to the general P1-LL backend to store results. The subjects in frame are analysed so that a skeleton model of their limbs is constructed, an elements such as the speed of movement of the limbs can be recorded and saved for sessions.

Acoustic and linguistic features analysis [UVIGO]

The audio recordings and analysis for the pilot feature no major changes from the description in D5.1 and D5.2. The recording system uses two types of microphones: a lapel mic and the built-in microphone of a tablet or other edge device. The lapel microphone operates wirelessly on the 2.4 GHz band, with a transmitter worn by the participant and a receiver connected via USB-C to the edge device. The recordings are managed by a web application that uses HTTPS to ensure secure and encrypted data transmission. For connectivity, the edge device is designed to use a high-speed Wi-Fi 6E network, but it can also be used with a wired Ethernet connection. The recording application worked as expected, and we successfully collected audio from all participants on-site without any issues.

The team has been working on acoustic analysis for this project using publicly available datasets to identify the best paralinguistic, acoustic, and linguistic parameters, as well as machine learning classifiers. Based on the state of the art and our own tests on these datasets, fluency tasks seem crucial for correctly assessing the participants. Additionally, read tasks have shown promising results that will be helpful once the participants' audio recordings are available. Preliminary studies using the participants' audio recordings show promising results for our ability to identify key characteristics. We expect these findings to be further validated with a larger group of participants in the next cycle.

Gait analysis [EVALAN]

The Gait Instability Score (GIS) is a dimensionless metric that quantifies irregularities in a person′s gait. EVALAN has developed a chest worn sensor to quantify the GIS using the input of a three-axes accelerometer sampled at 104 Hz. The algorithm used to determine the GIS has been developed by the US Army Research Institute of Environmental Medicine [Buller M. et al., Br J Sports Med 2022; 0:1-7] and has been used so far to determine gait instabilities in athletes and military personnel.

The EVALAN gateway stores both accelerometer information and GIS calculation, which indicates the person′s activity: walking, running, or other, based on the detection and classification of individual steps in time intervals of 5 seconds. Raw data and calculated GIS are monitored during the session with the EVALAN Wobble app and can be downloaded at the end of the session for offline analysis.

Gait and postural pattern analysis [CSIC]

The system described in previous deliverable documents for pilot P1-LL consists in two sensor subsystems:

Foot-mounted inertial measurement units (IMU), which collect accelerometer and gyroscope signals, for short periods of time (short walks), then process these streams to generate parameters related to gait quality: cadence, step length, floor clearance, and so on.

An RBG / depth-camera placed in the Living lab, capturing video streams for short periods of time (a few minutes) while patients perform a set of sample rehabilitation exercises in front of it. This is processed to extract parameters related to exercise performance, such as limb speed, joint angles, etc.

Both sensor systems have been evaluated successfully in the first stage with the reduced set of patients described in section 3.1.1 (both with impairments and a control group), and we are currently examining the results of the monitorization.

The sessions included 15 participants (13 with valid data). Each trial lasted between 8 and 19 minutes, with distances covered ranging from 183 m to 475 m. The number of strides analysed per subject varied between 40 and 160, which allowed us to obtain a representative sample of their gait pattern.

The following figure shows the results obtained by the CSIC in the first cycle experiments carried out at IRBLL Living Lab facilities

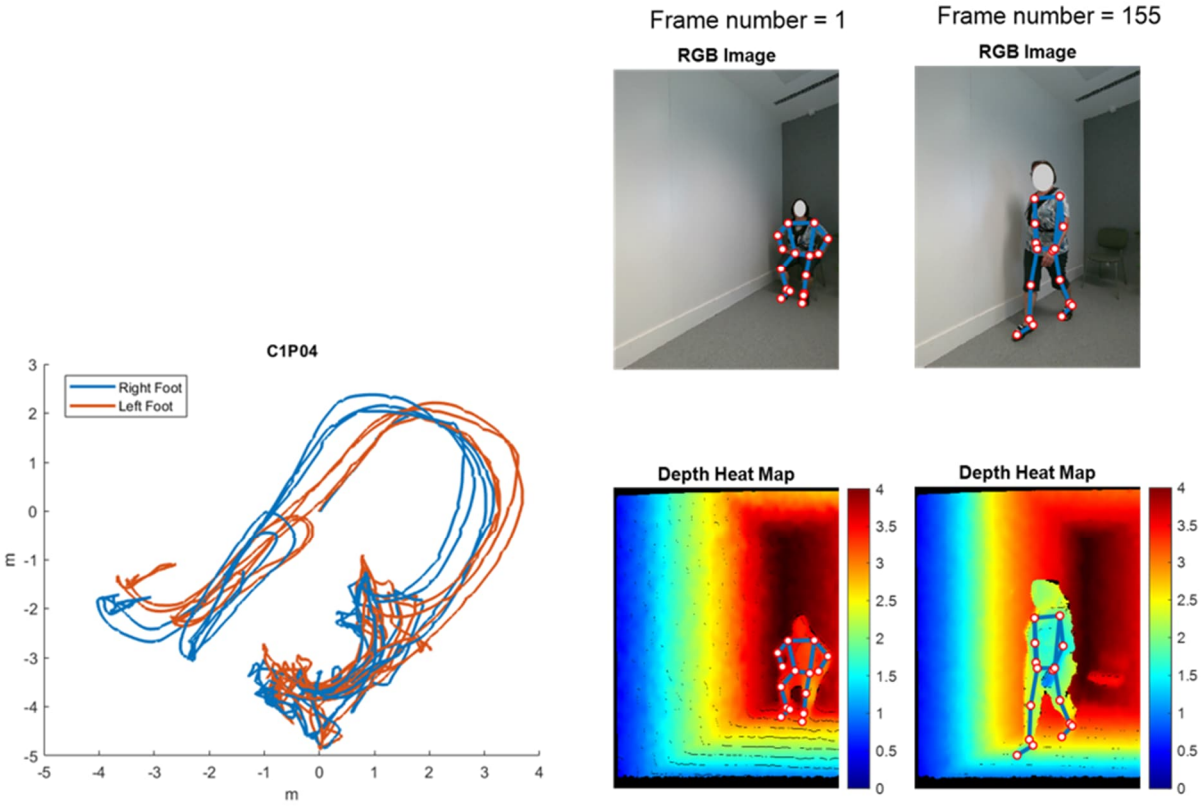


Figure 4 First cycle results of the CSIC sensors at the Living Lab in IRBLL. Left: IMU-estimated trajectories for both feet of a patient; right: monitorization of a Timed Up and Go (TUG) frailty test with the depth-camera

We have obtained feedback from the people operating the sensor systems, and we'll take it into account to redesign and simplify the user interfaces for the sensors for cycle 2.

Object identification and range estimation [ACORDE]

A multi-sensor monitoring platform has been implemented for the detection and localization of people and objects, based on an AWR1843 FMCW Evaluation Board radar and an RGB camera provided by HIB, which is also used as a complementary sensor. The AWR1843 radar generates a partially processed data array, which is sent to a gateway (GW). The GW transmits this array to the Edge device, where it is stored and mapped to determine the positions of detected objects and the number of elements present within the monitored area.

The RGB camera captures real-time images, which are processed using Deep Learning models to enable detection and localization of people and objects. This produces precise position maps, providing semantic information of the scene and high spatial resolution of the detected elements.

In the current phase of the project, the work is focused on validating each sensor independently for the detection and localization of people and objects. The final goal of project is the integration of both sensors, enabling data fusion to enhance robustness and accuracy, and extending the system's capabilities towards fall detection in people.

Blockchain-based authentication for IoMT [ODINS]

In addition to the multi-sensor system, P1-LL also incorporates a blockchain-based authentication framework into the Edge-Gateway. As previously articulated in earlier deliverables, the gateway was

engineered to function as a modular and adaptable hub linking diverse sensors and processing units. This new integration enhances the existing design by embedding a security layer directly into the gateway, guaranteeing that all connected devices undergo proper authentication prior to their data entering the system.

At a macro level, the procedure adheres to a simple flow. When a sensor or device establishes a connection with the gateway, it is initially identified and registered through the blockchain protocol. Subsequently, the gateway authenticates each device whenever it transmits data, permitting only verified devices to engage in the monitoring and decision-making processes. This indicates that the gateway not only consolidates and processes data from multiple sources but also enforces trust at the very initial point of interaction.

The advantage of this integration lies in its provision of a transparent yet robust security layer to the pilot, without modifying the manner in which participants or caregivers engage with the system. For end users, the system operates in the same manner, while in the background, the gateway ensures that all data entering the health monitoring model originates from trusted and verifiable sources. This is especially crucial in a pilot such as P1-LL, which encompasses sensitive medical information and the involvement of vulnerable populations.

During the subsequent cycle of the pilot, the blockchain-enabled gateway will be evaluated concurrently with the multi-sensor infrastructure, ensuring that the authentication mechanism does not impede usability, introduce latency, or adversely affect overall system performance.

3.1.3 Pilot technical KPI measurements

[UVIGO] No changes have been made to the KPIs from D5.2. While the first cycle of our work focused on the foundational data model and a lab-based proof of concept, the second cycle will concentrate on integrating these developments into the pilot program. Our objectives will be directly tied to the remaining KPI of MCI estimation via audio signal analysis (Determining key characteristics for detecting MCI). Specifically, we will work to determine the key characteristics for detecting MCI by testing and analyzing the features and models selected in the first cycle with participants' audio tasks. Our goal is to statistically identify the key characteristics for MCI.

During this first cycle, we tested our foundational data model with a few participants' audio recordings. Our preliminary analysis suggests that some characteristics appear to be statistically important for the determination of MCI. In the next cycle, we will validate these findings—and identify other useful characteristics—by testing them with a larger dataset composed exclusively of participants' data.

[ACORDE] In this work cycle, activities have focused on the detection of people and objects. For data acquisition, the FMCW radar was used in combination with the RGB camera provided by partner HIB, with tests carried out in the pilot of P1-LL involving different patients under controlled conditions.

The experimental process consisted of the synchronized capture of multichannel information from both sensors, followed by off-line processing using detection and classification algorithms. Subsequently, sensor data fusion was applied with the aim of leveraging the complementarity between technologies.

The results obtained confirm that the system is able to meet the KPIs committed for this first phase, both in terms of detection range and response time. In particular, measurements were successfully performed below the threshold of 10m, as defined in one of the KPIs, and the average

acquisition/processing speed was around 1 second. This achievement validates the proposed strategy and represents a key step toward the consolidation of the system.

Looking ahead, the next phases will focus on the migration of processing to an Edge-Gateway environment, enabling real-time data fusion and analysis, improving efficiency, and facilitating integration into practical applications.

[ODINS] In this project, ODINS seeks to evaluate in a practical manner a secure, lightweight authentication system for IoMT devices, leveraging the Hyperledger Fabric platform. These indicators have been established considering each component and functionality of the system, including the incorporation of security solutions for IoT, which often require high resources, making them unfeasible for IoMT networks with limited processing, storage, and battery capacity. Furthermore, traditional centralised approaches present single points of failure, increasing vulnerability to attacks such as DoS. Each indicator has a specific target value or qualitative goal that serves as a benchmark for verifying the progress and effectiveness of the solutions implemented. Consequently, the blockchain-based proposal seeks to overcome these limitations by offering an efficient security architecture that guarantees the protection of sensitive user data during transmission on the IoMT network. Hyperledger Fabric allows permissioned blockchains to be deployed without the need for costly consensus protocols such as Proof of Work, making it suitable for IoT-based health monitoring systems. Therefore, to assess the performance of authentication mechanisms, indicators such as energy consumption per registration request (approximately 60 mW), the time required to generate messages (~30 ms), the blockchain's capacity to process transactions (~800 per minute) and the execution time of each transaction (~40 ms) will be taken into account. The goal is to achieve an approximate 10% improvement in each case, demonstrating that the solution is efficient, robust, and suitable for resource-constrained IoMT environments. An overview of the KPIs can be found in Table 2.

Figure 5 shows the distribution of response times in the identity registration process within the blockchain network. Through the average and different percentiles (90, 95, and 99), it is possible to observe the general trend of the system underload and the existence of variations in latency. This information is useful for identifying both the usual behavior of the authentication mechanism and the less frequent cases in which longer delays occur, offering a more complete view of the stability and efficiency of the system.

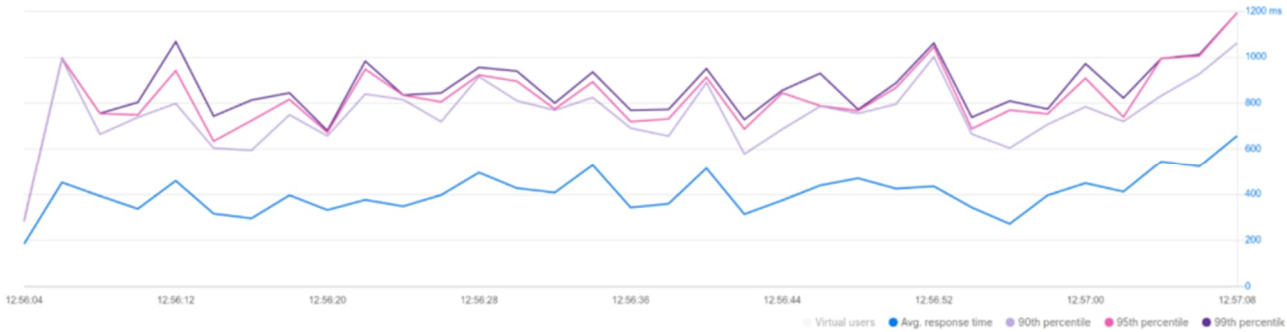


Figure 5 Response and processing time metrics of the proposed blockchain-based authentication mechanism (local deployment).

Figure 6 shows the system's throughput during the same test, in which 813 registration requests were executed in a 60-second interval. This equates to an approximate throughput of 13.55 requests per second, reflecting the ability of the blockchain-based authentication mechanism to continuously

handle a sustained volume of transactions on IoMT networks. The result shows that the solution maintains a constant flow of error-free registrations, ensuring both the stability and reliability of the authentication process.

Figure 6 Throughput performance of the blockchain-based authentication mechanism (transactions per minute).

All results presented correspond to tests carried out in a local environment, which means that they still represent a controlled and preliminary scenario. Both the KPIs already measured and those still to be evaluated will be analysed again once the blockchain network is fully integrated into the pilot, allowing performance to be validated in a real IoMT context.

Table 2 P1-LL Technical KPIs

KPI	Validation Metric	Status
O3 KPIs		
3.1.2 Integration of multiple sensor flows from heterogeneous communication interfaces	1. Abstraction layer for data collection from wired and wireless communication interfaces	1. Proof of concept deployment achieved collecting data from four sensor technologies.
3.2.1 Traffic transmission in the system	1. Data is transferred directly from the source to the destination to consume it.	1. First prototype achieved with one connection using the overall DistriMuSe platform (WP4)
3.2.4 Orchestration of applications	1. Launching time of an aggregated deployment in the same network.	1. Proof of concept for most sensors, one usage of the distributed platform and orchestrator achieved.
3.5 Blockchain-based authentication for Internet of Medical Things (IoMT)	1. The implemented blockchain-based lightweight authentication mechanisms will be evaluated based on their i) computational cost to support the authentication mechanism, ii) latency, iii) throughput and iv) processing time required to execute transactions	i) average power consumption for registration request generation: around 60 mW; ii) average time to generate messages and registration request: around 30

		ms; iii) average throughput: 800 transactions/min; iv) average processing time: 40 ms
O4 KPIs		
4.1 Continuous remote and unobtrusive monitoring of physiological parameters	1. Physiological parameters measurements while performing activities	1. Proof of concept achieved for the chosen parameters,
4.3 Position and pose extraction	Tracking of user trajectories in the living lab, with segmentation into steps, and computation of gait quality parameters Identification of enough body features to allow assessment of human motion when performing a restricted set of exercises in the living lab	Both capabilities were demonstrated with the group selected for first cycle trials, although for some subjects we didn't get valid measurements
4.4 Activity detection	1. Successful monitoring of daily activity, detection of changes in activity 2. Correlation between remotely detected movement patterns (camera-based) and wearable sensor-based data	Will be addressed in the second cycle of the pilot evaluation, once we can perform longer data acquisition campaigns.
MCI estimation via audio signal analysis	1. Determining key characteristics for detecting MCI	1. Preliminary analysis of audio recordings shows promising results, which will be validated in Cycle 2.
O5 KPIs		
UC1 MCI tracking & lifestyle monitoring	1. Analysis of initial patient interviews 2. Daily Activity recognition monitoring 3. MCI index calculation to explain the condition of a patient	1. Preliminary analysis complete

3.1.4 User aspects: stakeholder engagement in pilot development

During the first cycle of P1-LL, user experiences were collected from patient study participants to identify areas for improvement for the second cycle and ensure that the solution aligns with their needs. User experiences were collected from 15 participants. Stakeholder perspectives were collected through a validated questionnaire and in individual semi-structured interviews. Both instruments were administered to participants after they had completed the study activities detailed in Section 3.1.2.

The System Usability Scale was employed to measure participants' perceptions of the usability of the multi-sensor system as a whole. The System Usability Scale is a validated and one of the most widely used tools to measure individuals' perceived usability of technology.

Semi-structured interviews consisted of questions designed to gain a deeper understanding of user experiences. Interviews focused on patients' previous experiences with health monitoring

technology, their perceived barriers and facilitators to use, acceptance of health monitoring technology and intention to use, as well as the supports they require to engage with health monitoring technology in the future confidently.

Preliminary findings from the System Usability Scale indicate that study participants rated the usability of health monitoring technology ranging from marginal to acceptable. Although the majority of participants indicated willingness to use health monitoring technology, the need to learn a lot of things and the necessity of technical support were identified as potential barriers.

The in-depth interviews further support the quantitative findings. The following five themes were identified in the preliminary analysis of the semi-structured interviews with study participants:

1. **Familiarity with health monitoring technologies**: Participants showed limited prior exposure to health monitoring technologies. Their main reference point was the emergency call button, widely recognized as a device for safety in the event of a fall. Smartphones, wearables, and continuous monitoring systems were not generally understood as health-related tools.
2. **Prevention versus emergency response**: Most participants associated health monitoring technologies with monitoring for emergencies rather than prevention. Even after discussion about early detection and preventative care, participants questioned the usefulness of technology, particularly those who considered themselves “healthy.” Instead, monitoring was perceived as more appropriate for those with cognitive decline or established illness.
3. **Privacy, safety, and feelings of control**: Views on privacy were mixed. For some, cameras and monitoring systems were seen as intrusive, while others prioritized safety and argued that surveillance was acceptable if it ensured protection and reduced risks. Importantly, participants also reflected on the possibility of feeling controlled in their own homes. However, when asked about the observation during testing, many reported that they would forget about the sensors and wearables during the tests.
4. **Technical support**: Several participants anticipated difficulties in managing and operating the technology independently at home, expressing concerns regarding understanding how devices work, setting them up, and troubleshooting problems. For this reason, they emphasized the importance of having external support, such as trained personnel or technicians, who could install the technology, explain its functions, and provide ongoing assistance if needed.
5. **Access to information**: Most participants believed that health professionals should be the main recipients of monitoring data, while some also considered it useful for family members to have access. At the same time, one participant preferred not to receive any information personally, as it could generate anxiety.

3.1.5 User-based KPI assessment

Some user-based KPIs for cycle 1 have been measured; however, the analysis is ongoing.

KPI	Validation Metric	Status
Participant recruitment	1. Recruitment target for each sub-group: MCI (n=5), PD (n=5), control (n=5).	1. Complete 2. Not started

	2. Recruitment of healthcare professionals for focus groups (n=10).	
System usability	1. System usability measured with patient participants	1. Complete
User experience	1. In-depth interviews with study participants	1. Complete

3.1.6 User aspects: Gender/age issues and ethical concerns in P1-LL

An ethical evaluation focus group was conducted to discuss the ethical implications of using home monitoring technology for older adults with neurodegenerative disease. Thirteen individuals participated in the P1-LL ethical evaluation focus group, eight women and five men, all directly involved in the development, management, or execution of the P1-LL pilot. Their roles included technology developers and coordinators, R&D project managers, researchers, scientists, and clinicians. The combination of technological and clinical partners informed a balanced discussion.

Informed consent was collected from participants before the focus group and reinforced in the meeting orally prior to the discussion beginning. The focus group was conducted using Microsoft Teams, was audio-recorded, and lasted approximately one and a half hours. The discussion was facilitated using the provided presentation template by the partner in charge of compiling the ethics exercise form (IRBLL). The discussion covered four themes: anticipation, reflexivity, inclusiveness, and responsiveness. Following the focus group, the partner in charge synthesized the responses and shared them with the focus group participants to ensure coherence. The final version of responses was submitted to the ethics exercise tool and will be presented in deliverable D7.7.

3.1.7 P1-LL transition into cycle 2: takeaways and feedback

During Cycle 1, all components of the sensor systems were trialed with study participants. While all monitorization sensors operated correctly, there were some technical difficulties related to sensor setups with some participants, connectivity issues, and difficulties operating the provided software, all of which we deem to be expected when working with new technology. We will address these problems in Cycle 2 of the pilot P1-LL. Based on the preliminary findings of user usability from Cycle 1, a further mixed methods evaluation of user experiences with a larger sample is planned for Cycle 2, including engagement with health care professionals and care givers of older adults with neurological diseases.

3.2 P1-KUO evaluation cycle 1

P1-KUO is delayed, due to late national funding decisions, influencing the Finnish partners. There are two subpilots: monitoring the elderly in Kuopio (led by VTT) and the satellite pilot led by Benete in Helsinki, focusing on creating automated patient reported outcomes for the elderly population. In the first stage the impact of muscle training will be automatically evaluated.

The Kuopio pilot has not started, we have a positive medical ethical statement but we still need to acquire a permission for clinical investigation from The Finnish Medicines Agency Fimea, as the VTT radar is considered a ”non-CE marked medical device” https://fimea.fi/en/medical-devices/investigations-with-devices/clinical-evaluation-and-clinical-investigations. However, we can start implementation and recruitment yet.

The Helsinki pilot has started with general planning and installations.

3.2.1 Final pilot set-up

3.2.1.1 Kuopio pilot

For the Kuopio pilot, the implementation is in starting phase and the final pilot setup is in progress. The set-up is summarized in 3.2.2

We have revised the research plan according to comments from the Medical Ethical board of Kuopio University Hospital. The positive decision was received 23.8.2025. The alterations were related to informed consent collection process and to ensuring that the home-monitoring data does not contain any data from other people, such as visitors.

3.2.1.2 Helsinki subpilot (HEL sub)

The Helsinki pilot environment was established in early April 2025. The installation covers five apartments in a nursing home. In line with the research plan, the first phase of the project has been completed. This included setting up the technical environment for data transmission and storage (base stations, network connections, servers, etc.). Norlandia has selected the participating residents and obtained their consent. Development of the algorithms required for producing the AROMs has also been initiated (Active Range of Motion).

3.2.2 Pilot evaluation execution and protocol details

The Kuopio pilot has two monitoring contexts: a three-week home monitoring phase, and a controlled test session.

In-apartment monitoring: The monitoring technology is embedded into a chair and a bed sensor in the apartments of the residents. A pressure-sensing seat foil is used to detect the pressure distribution, breathing rate and pulse. The bed sensors are inserted underneath the mattress to measure pulse, breathing, movements and time in bed. The sensor is also used to collect the audio band (kHz region) movements, which provides information about sleep sounds. For 1-2 nights, a pulse oximeter is added to the setup, to augment the sleep data and to enable interpretation of sleep data in terms of sleep laboratory (P1-KSL pilot) expert domain.

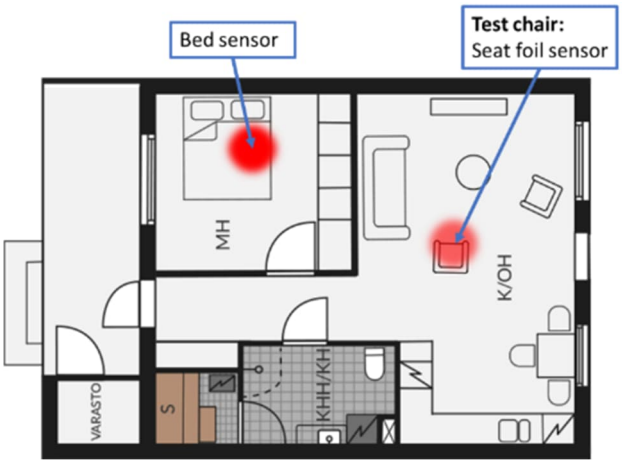


Figure 7 The setup for non-contact monitoring in the apartment

In the Helsinki subpilot the monitoring technology consists of

- Living room motion sensor to measure presence and activity
- Bathroom motion sensor to measure presence and activity
- Bedside sensor for detecting resident getting out of bed
- Toilet sensors to detect time and duration of toilet visits
- Shower sensors to detect time and duration of shower visits
- Door sensor for detecting door usage
- Sleep sensor (commercially available medical device) to measure time in bed, sleep interruptions, sleep qualities, pulse, respiratory rate and sleep apnea
- Walker sensor to measure when and how often the walking aid is needed
- Gateway (base station) for transmitting the data to Benete's backend solution

Controlled tests: Frailty and general functional capability will be assessed in a separate controlled study session, to which participants are invited one at the time. This information is complementary to the frailty index computed from InterRAI-HC patient record data (Resident Assessment Instruments for Home Care). We will investigate how the InterRAI-based frailty index correlates to the functional capability results as well as to the behavioural (movement) data collected by monitoring.

The technology development and validation goals are mostly related to the VTT radar. We will use accelerometers, heart rate sensors and video as reference and investigate the performance of radar data in assessment of mobility, vital signs and frailty. The data will enable development and optimization of radar signal algorithms on different levels (raw data – feature extraction – modelling).

The functional capability tasks in the frailty assessment are:

- 2-minute baseline (sitting in the chair)
- Rise from the chair five times without using hands (if capable)
- 2-minute recovery (sitting in the chair)
- Tandem stance (short physical performance battery)
- Timed-up-and-go (TUG)
- 2-minute recovery (sitting in the chair)
- Walking speed protocol: the person is asked to walk for 10 m and the time used is measured.
- 2-minute recovery (sitting in the chair)

- Grip strength measurement, using a standardized handle

In total the frailty test protocol will last ca. 25 minutes, depending on the participant's performance.

We will use standard agility and frailty assessment data RAI assessments[2], of which InterRAI Home care (InterRAI-HC) is used for older adults living at home. The RAI data will be accessed by the nursing staff, and they will compute the frailty index using a spreadsheet provided by the researchers. Only the index will be reported to the researchers. We will also interview the participants during the home-monitoring phase and at the controlled sessions as well as collect background information using questionnaires.

The cross-pollination within DistriMuSe will focus on sleep data. We are closely collaborating with P1-KSL sleep study, and aim to compare the in-apartment sleep data with the golden standard sleep laboratory PSG data, in order to 1) understand the validity of sleep data acquired using the technology of P1-KUO, and 2) increase the understanding of the sleep of the elderly who are not investigated for a particular sleep problem. In addition to sleep analysis, to enrich the data analysis and interpretation of movement data, experience exchange and possibly sharing of algorithms will be sought with other DistriMuSe pilots working on movement analysis

The key core DistriMuSe technology used in P1-KUO is the polarimetric VTT radar (WP2).

3.2.3 Pilot technical KPI measurements

Table 3 Technical KPIs for P1-KUO. As the main pilot in Kuopio is delayed and not started yet, the status of the KPIs is shown only for the Helsinki subpilot lead by Benete.

KPI	Validation Metric	Status (subpilot)
Implementation of the multisensory continuous *in-apartment monitoring* environment Implementation of the multisensory monitoring environment for the *controlled tests*	1. All sensors successfully implemented and tested 2. Data connectivity architecture ready and tested 3. Data synchronization at required level: tolerances: max. 60 seconds for continuous monitoring and 500 ms for controlled tests	1. Done 2. Done 3. Done
Conversion of fused sensor data into actionable information of health status (related to 4.6 Health assessment based on fusion of data)	1. Algorithms and model(s) for comprehensive health status created: physiological status, behavioural status, sleep status 2. Indices for feature relevant for caregivers created 3. Indices and thresholds for abnormality alerts created	1. Started 2. Started 3. Started
O4 KPIs that are directly relevant to the pilot		
4.1 Continuous remote and unobtrusive monitoring of physiological parameters	1. Breathing and heart rate reliably detected during sleep, using the bed sensor 2. Breathing and heart rate reliably detected during day, while seated, using the seat foil sensor	1. n/a 2. n/a
4.2 Mental state extraction	3. Changes in stress levels, sleep problems and daily pattern abnormalities detected, using participant and caregiver reports as ground truth	1. Started

[2] https://thl.fi/en/topics/ageing/assessment-of-service-needs-with-the-rai-system/information-on-the-rai-assessment-system - Information on the RAI Assessment system

4.4 Activity detection	1. During home monitoring: successful monitoring of daily activity, detection of changes in activity 2. During controlled sessions: Correlation between remotely detected movement patterns (radar, camera-based) and wearable sensor-based data 3. During controlled sessions: Correlation between remotely detected movement patterns (radar, camera-based) and routine functionality assessment results	1. Started 2. n/a 3. Not started
O1 and O2 KPIs that are relevant to the pilot as enabler		
1.1 Radar-based sensing for physiological signal monitoring	Utilization of improved detection of vital signs: robustness, accuracy and algorithms improved significantly above the status quo	1. n/a
1.2 Radar-based sensing for human motion detection	Utilization of improved detection accuracy: 90 % of body movements detected	1. n/a
1.4 Camera-based sensing	Utilization of improved specificity and sensitivity	1. n/a
1.6 Positioning of people in indoor environments	Utilization of improved specificity and sensitivity	1. n/a
2.2 Late fusion	AI-based fusion of different signal sources for computation of health-related indices	1. Started

No changes in from D5.1/D5.2 technical KPIs are made. Evaluation of the technical KPIs is not possible at this stage.

3.2.4 User aspects: stakeholder engagement in pilot development

The main stakeholders are involved in the planning. The pilot (both Kuopio and Helsinki) use Norlandia Care facilities as research site, and the personnel participate in planning, implementation and also actual data collection. Another key stakeholder group are monitoring technology companies. In the pilot, both Benete and eLive are actively contributing and processing the pilot outcome. The end users, the elderly residents, are a crucial stakeholder group. Their views and opinions on the technology, acceptance of home monitoring, and data privacy/sharing are carefully researched and disseminated to other stakeholders and DistriMuSe community.

The DistriMuSe ethical process has already addressed the pilot in a focus group interview. The purpose of these interviews is to investigate the pilot implementation from the end-user and service provider perspective as well as to discuss ethical aspects of the future service and product. In the discussion among the main stakeholders, following points were raised:

- In-apartment monitoring technology aims in building trust and security. The health problems will be recognized at early stage, even in advance, and the right actions can be taken swiftly. Continuous monitoring will improve the understanding of the status and help to recognize changes. This benefits not only the user but also the caregivers. Data-based decisions can be made.
- However, there is a danger of over-treatment, when there is more health information available. The balance should be carefully considered when the system is in use.
- One key question is what kind of information is needed to better anticipate care work, and how this anticipatory information is used. In the real world, the details of technology is not the point. Technology is a spade, and it is digging the ditch that matters, not the design of

the spade – as long as it works. This should lead the research, and also the communication to users and other stakeholders.

- Adaptation of novel technology requires resources, both training and in everyday work. How to communicate the solution to relatives and medical staff. How are the relevant people (e.g. different members of medical and nursing care providers) and systems be recognized and involved? The flow of monitoring data in the organization must be planned in detail: to ensure all who need the information get it in the right scope and format, while preserving the privacy.
- Diversity of the users was discussed. Language skills, ethnical and cultural background will vary even more in the future. These must be considered in all solutions and communication.
- The ethical stress of the personnel was brought up: how to balance with observed indicators of health problems vs. restrictions in care or limited rights to report the situation?

3.2.5 User-based KPI assessment

Regarding the stage of the pilot, most KPIs cannot be addressed yet.

Table 4 User-based KPIs. The indicators that have been reached are marked in bold. As the main pilot in Kuopio is delayed and not started yet, the status of the KPIs is shown only for the Helsinki subpilot lead by Benete.

KPI	Validation Metric	Status (subpilot)
The system is adequately tested	**1. The test users represent the focus group** 2. N = 25 is reached 3. Ground truth data is valid and relevant to assess the health status 4. The models and cross-validation are done carefully to minimize biases and to ensure best possible generalizability of the results	1. Started 2. Not reached 3. Started 4. Not started
The system provides useful information of the person's status	1. Validation of the data against routine health data 2. Discussion about the relevance of the data with end users (elderly people, caregivers) 3. Fine-tuning of the health status indicators and anomaly alerts based on the insights of the professionals	1. Started 2. Started 3. Not started
The system is considered acceptable and ethically justified	**1. The pilot plans are put under the ethical evaluation process of the local University Hospital** 2. The setup and information it will yield is discussed in focus group interviews (participants and other stakeholders), as well as informally during the collaboration with health care professionals	1. n/a 2. Started

3.2.6 User aspects: Gender/age issues and ethical concerns in P1-KUO

The pilot is included in the *DistriMuSe Ethics Exercise Tool*. As mentioned above, the focus group interview was conducted. In addition, DistriMuSe ethical advisor provided assistance in collecting the ethical evaluation from the Kuopio University Hospital Ethical Board.

In the recruitment, the process of selecting candidates and particularly on evaluation of their capability to give informed consent, was refined based on Ethical Board comments. We will enroll only persons who can give consent. The nursing staff evaluation will now be confirmed by the medical doctor who knowns the candidate. In case of doubt, the person will not be evaluated more

closely, but instead, the recruitment of that person will not be continued. Another concern from the Ethical Board was that the sensors are insensitive to the person being monitored and data of a wrong person may be collected. To eliminate the chances of that, only persons living alone will be recruited. The chance that a visitor will use the sensored furniture will however remain. This risk will be lowered by explaining that the sensored chair and bed should only be used by the participant but the risk cannot be eliminated completely. Based on the previous study, the participants understand and follow these instructions very well.

3.2.7 P1-KUO transition into cycle 2: takeaways and feedback

At this stage, cycle 1 is delayed and hence cycle 2 takeaways are not yet available.

3.3 P1-G evaluation cycle 1

Pilot P1-G, coordinated by Televic in collaboration with UGent, KUL, Sentigrate, and COMmeto, focuses on validating lifestyle monitoring solutions in care facilities through a multi-sensor setup combined with advanced AI algorithms. The goal is to provide caregivers with intuitive insights into patients' daily activities and health conditions, enabling more effective responses to emergencies, improving follow-up care, and enhancing overall quality of life. The solution integrates radar-based sensors (FMCW and UWB), BLE/ultrasound localization tags and beacons, and personal alarming to monitor human activities and movements in a non-intrusive way. These inputs are processed through a layered architecture including sensor fusion, alarm generation, and device management components.

The primary users are caregivers and patients in residential care settings. The pilot emphasizes both technical objectives (accuracy and generalization of AI algorithms, feasibility of cost-effective sensor networks, and reliable communication protocols) and user-related objectives such as caregiver acceptance, usability, and integration into care workflows.

3.3.1 Final pilot set-up

In cycle 1, individual components were validated separately, including radar-based activity detection, BLE channel sounding for localization, Joint Communication And Sensing (JCAS) protocol development, sensor fusion, and device management and orchestration. Each component has its own KPIs and validation plan but were tested within the same living lab setting (UGent HomeLab). In cycle 2, these components will be integrated into a unified system with improved robustness, scalability, and user interaction tools.

In D5.1, a setup was described for the first two data campaigns, one with collaborative researchers and one with external participants. The data from these campaigns are used as training data for the algorithms. The data campaigns consisted of participants executing a sequence of defined activities in six different room layouts. To assess generalizability, a separate evaluation setup was created in the same room, featuring new room layouts, a different sequence of activities, and new participants. The pilot setup for the cycle 1 evaluation was previously described in D5.2 and is briefly repeated below for convenience in Figure 8.

The setup simulates a room within a care facility. The bedroom includes a private bathroom and is equipped with radar sensors to monitor various activities. The system is designed to detect movement, such as a person entering or leaving the room, either through the corridor or into the bathroom. These events are captured by both the radar system and an indoor positioning system. As illustrated in Figure 8, FMCW and UWB radars are mounted above the bed to detect in-room activity. Additionally, each room features a Televic AQURA Beacon, which enables room-level localization of patients wearing a Tag. The Tags also provide a personal alarm function through a built-in button. Alarms are received by AQURA Communicators, connected to the AQURA server. To enable within room-level localization using BLE HADM (High Accuracy Distance Measurement) the AQURA Beacons have been extended with nRF54L15 development kits. These kits support Bluetooth 6.0, which is required for BLE HADM functionality. The nRF54L15 will eventually be integrated into both the Beacons, Tags, and Communicators in cycle 2.

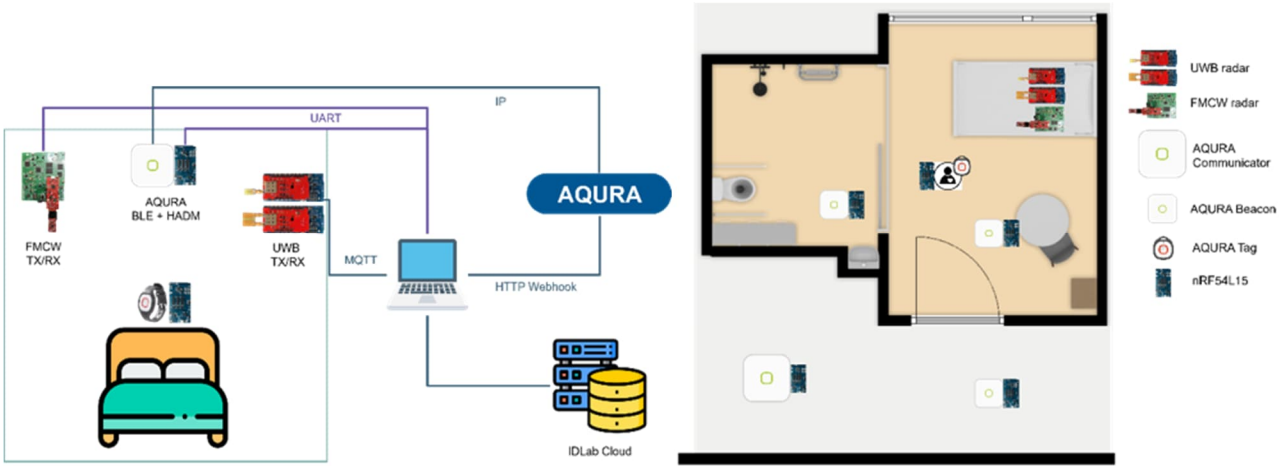


Figure 8 Final P1-G pilot setup

The new room layouts, differing from the layouts in D5.1 for the data campaigns, are shown below in Figure 9.

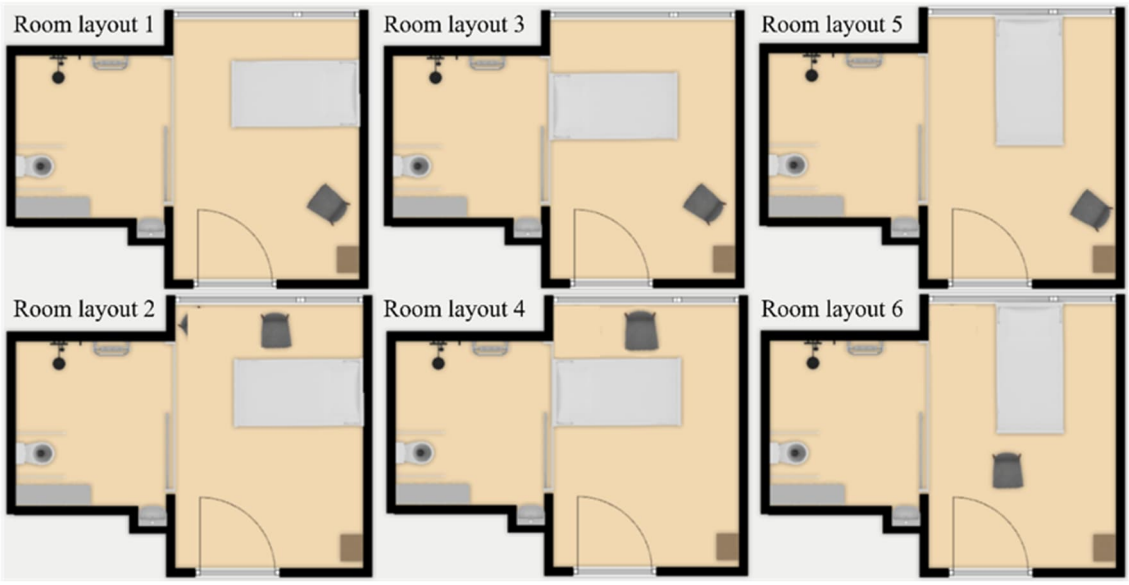


Figure 9 New room layouts for P1-G pilot evaluation

The sequence of activities, used in the pilot evaluation, are listed in Table 5. Unseen activities (not used in the data campaigns) are highlighted in grey. The sequence of activities was repeated for each room layout and six participants, of which four were unseen and two seen.

Table 5: Sequence of activities for P1-G pilot evaluation.

Fine Grained Activity	Coarse Grained Activity
Walk inside the room	Entering room
Wander around in the room	Walking
Walk to the bathroom and enter it	Walking to bathroom
Go back inside the room	Entering room
Walk towards the chair	Walking
Sit down on the chair	Sitting
Eating hand movements	Sitting
Stand up from the chair	Standing up
Go to the left side of the bed	Walking
Sit down on the edge of the bed	Sitting
Lay down on the bed	Laying down
Anxious, frequent movement in bed	Laying down
Press button on wearable	Laying down
Get up on the edge of the bed	Sitting
Lay down on the bed	Laying down
Stand up from the bed	Standing up
Go to the right side of the bed	Walking
Wave your hands	Standing up
Clap your hands	Standing up
Pick up coin from the floor	Standing up
Walk sideways out of the room	Walking
Wander around in the corridor	Walking

In order to evaluate the accuracy of BLE localization using channel sounding, each participant followed an indicated trajectory, shown in Figure 10, and remained stationary for 10 seconds at each known position indicated with an X.

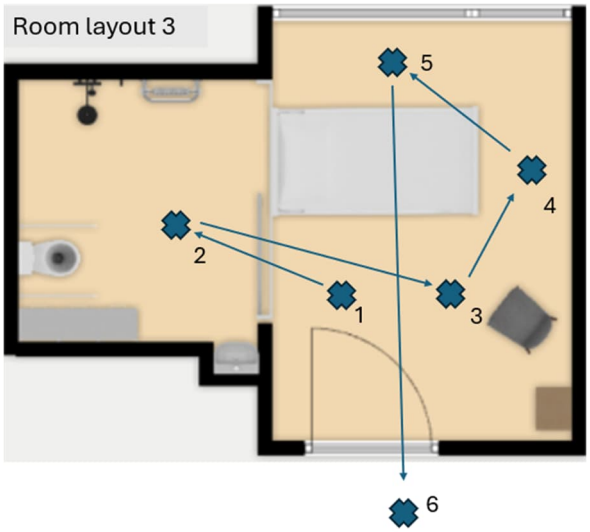


Figure 10 Trajectory for participants to evaluate BLE localization

3.3.2 Pilot evaluation execution and protocol details

The pilot architecture features no major changes compared to the description in D5.2 and D5.1. The architecture and the different components that are evaluated in cycle 1 are briefly repeated.

Four different sensors are deployed in the living lab: an FMCW radar, a UWB radar, a person localization system using BLE and ultrasound tags with beacons, and a Mobotix camera with integrated AI from Kepler. These sensors provide complementary data streams on presence, movement, and activities. On top of this, a sensor network infrastructure is in place to handle sensing, localization, and communication. A dedicated JCAS protocol stack is deployed to connect the sensor nodes with the gateway layer. The AQURA gateway layer takes responsibility for edge processing of the raw sensor data, ensuring that only meaningful and preprocessed information flows to the next layer. The AQURA processing layer integrates various AI algorithms. This includes a movement classifier that detects low-level activities from radar data, room-level localization using radar (without tags), localization based on BLE HADM with tags, and person-specific alarm triggers. Above this, the AQURA business logic layer is designed to implement software modules for creating alarms, workflows, and dashboards, which allow interaction with end users. Although this component is out of scope for the first cycle, it will be accessible through APIs to different applications in later phases. Finally, a device management and orchestration layer is included to design and apply device management concepts on edge devices in line with TR-369. This ensures secure, scalable, and efficient operation of all deployed devices.

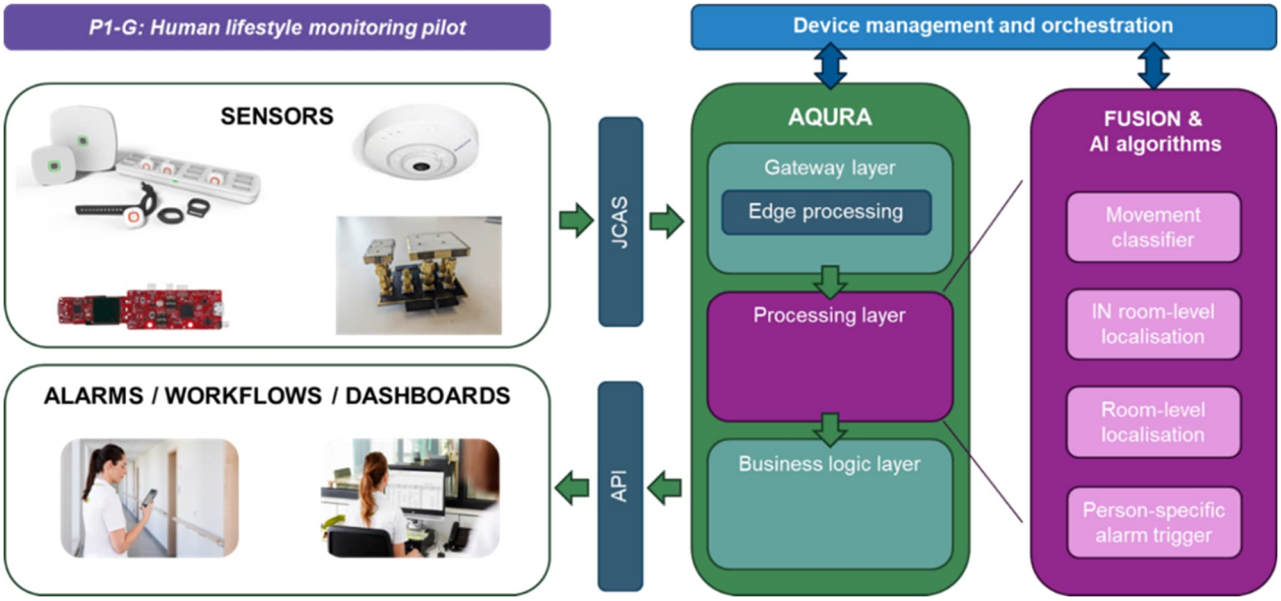


Figure 11 P1-G pilot architecture

The pilot has been split up into five components for the first cycle, where each component has its own KPIs and validation plan:

- Component 1: radar-based activity detection with UWB and FMCW
- Component 2: person localization using BLE channel sounding
- Component 3: joint communication and sensing protocol stack
- Component 4: sensor-fusion framework
- Component 5: device management and orchestration

The first component performs **radar-based activity detection with UWB and FMCW**. Both technologies are developed in parallel, and in cycle 2 they will be compared considering performance, robustness, energy consumption, and cost.

FMCW radar currently classifies six coarse-grained activities (walking, sitting down, lying down, standing up, getting up, hand movements) using time-Doppler maps processed by a CNN architecture. The system addresses the critical domain shift problem - where models trained in one environment perform poorly in new settings - through Domain Adversarial Neural Networks (DANN), which uses adversarial training to learn domain-invariant features via three components: a feature extractor, activity classifier, and domain discriminator that work together in a min-max optimization strategy. The CNN processes time-Doppler representations through four convolutional layers with increasing complexity, followed by fully connected layers for six-class classification. Data augmentation techniques including time-frequency warping and occlusion simulation enhance robustness by simulating movement variations and partial body occlusions. The system was trained on hospital and assisted living environment data, then evaluated in the HomeLab facility to demonstrate cross-domain generalization capabilities.

UWB radar is well suited for short-range monitoring and multipath-rich indoor environments. It relies on analyzing channel impulse responses (CIRs), which capture how UWB signals propagate and reflect in a room. Pre-processing steps such as phase normalization and clutter removal are used to highlight variations caused by human presence and movement. From these signals, range-time and time-Doppler representations are derived, which form the input for activity recognition algorithms.

The activity recognition is performed using convolutional neural networks (CNNs), which learn discriminative motion patterns from the radar-derived maps. At this stage, the focus is on sleep-related activities such as anxious bed movements, sleep interruption events, and nocturnal wandering, as these are best captured by a ceiling-mounted radar above the bed. Later experimental setups will extend the monitoring with additional receivers placed in the corners of the room to enable room-level activity detection.

The second component enables **person localization using BLE channel sounding**, a new feature introduced in Bluetooth Core Specification 6.0, by developing effective algorithms for indoor distance estimation and localization. Bluetooth Channel Sounding is centered on Phase-Based Ranging (PBR), which measures the phase shift of radio signals transmitted across multiple frequencies. By combining these measurements, the system can estimate the actual distance between devices. With multiple fixed anchor nodes placed throughout a resident's room, bathroom, and corridor, PBR-based ranging can be used to estimate the distance to each anchor and the patient's position through multilateration.

Several distance estimation algorithms have been evaluated using the same dataset. These include two slope-based methods, an FFT-based method, and multiple MUSIC-based super-resolution approaches. The MUSIC algorithm achieved the most accurate results and is, therefore, used during the pilot evaluation. For localization, the ranging estimates are combined through trilateration, which computes the patient's position from the distances to multiple fixed anchors. The evaluation considers both point-level accuracy (precise coordinates within a space) and room-level accuracy (correctly identifying whether the patient is in the bedroom, bathroom, or corridor).

The **Joint Communication And Sensing (JCAS) stack** is responsible for supporting communication, localization, and sensing with various data rates, application requirements, and network topologies. The pilot setup for cycle one features a Televic AQURA localization and personal alarming system that was adapted to use BLE communication instead of a current proprietary RF protocol at 868 MHz. This enables the communication part of the JCAS stack. The operation of the AQURA system and relevant communication of the JCAS stack was previously described in D2.6 and D2.7, but is summarized below for convenience.

The AQURA system uses ultrasonic waves to localize patients wearing a Tag (US receivers), which can also be used for personal alarming. Beacons transmit US packets so a patient Tag knows which room it is located in by listening to the transmitted US packets. Upon a location update, it broadcasts a location update message to nearby BLE Gateways using extended advertisements. These BLE Gateways are responsible to synchronize Beacons, as they use a TDMA scheme to prevent US collisions with nearby Beacons. In addition, patient Tags broadcast personal alarms upon a button press to nearby BLE Gateways, using extended advertisements with an application-layer acknowledgement to increase reliability. As shown in Figure 8, the pilot setup features one BLE Gateway in the hallway, three Beacons (one in each room) and one patient Tag.

Within room-level localization and sensing were not yet integrated into the pilot setup of the JCAS stack. Instead, within room-level localization using BLE channel sounding was evaluated separately using different devices. In addition, as no preprocessing or algorithms were deployed on embedded sensors yet, the distribution of sensor data was not taken into account. However, the available throughput within the JCAS stack, given application requirements, was evaluated previously in D2.7.

The **sensor fusion framework** combines information from multiple sensors: FMCW radar, UWB radar, BLE-HADM localization (within-room), and Televic localization (room-level), into a

representation of human activity and localization. By integrating heterogeneous data streams, the system becomes more resilient to individual sensor outages, noisy measurements, and inconsistencies between modalities. The goal is not only to improve activity recognition and localization accuracy, but also to detect unexpected behavior in the data, identify faulty sensors, and raise alarms related to either technical failures or patient monitoring events. The system should be extensible to adding new sensor modalities and robust to sensor outages or failures.

The development focuses on the sensor fusion algorithm on two levels.

- Early fusion: with a current focus on UWB and FMCW fusion for activity classification. Here the aim is to perform sensor fusion on the low-level representations of data (e.g. range doppler maps) from both sources. We will experiment primarily with representation learning of individual sources. One such approach is for example, training individual models on FMCW and UWB, then removing the classification head from both models. After which we can train a join classification head on top of the concatenation of the learnt representations.
- Late fusion: here the aim is to fuse high level algorithm decisions (e.g. decision of individual models) in hopes to collectively obtain a better decision. Primary methods consists of Kalman Filtering and weighted voting techniques.

The **device management and orchestration** component provides a robust and secure framework for remotely configuring, monitoring, and coordinating diverse devices within the lifestyle monitoring environment. Built on the TR-369 User Services Platform (USP), it enables integration of sensors and edge devices, supporting tasks such as parameter configuration, firmware updates, data collection, and software deployment. A particular focus lies on distributing AI-based processing modules to edge nodes where real-time analysis is required, while ensuring secure communication and protecting patient privacy.

During cycle 1, an extension to the USP data model (TR-369) was created, which was proposed to the Broadband Forum (BBF) consortium for inclusion in the upcoming revision of the standard (2.21). These components are organized under the IoTDataPathConfig section of the model. This model, structured across four technology-dependent layers and one technology-independent layer, provides a standardized way to manage and process sensor data at the gateway.

- Transport layer (e.g., MQTT, AMQP, HTTP, I2C, one-wire, etc.)
- Session Layer: describes associated sessions (ConsumerSession, ProducerSession). on top of the Transport layer
- The Modulation Layer: specifies how data is represented and encapsulated.
- The Profile Layer: describes how sensor and actuator values are mapped into standardized IoTCapabilities (Sensor and Control).
- The Binding Layer: maps technology-independent IoTCapabilities across technologies, enabling full cross-technology gatewaying.

This approach provides a robust framework for cross-technology gatewaying, allowing any sensor or actor to be represented and managed independently of its underlying technology. This not only supports individual sensors and actors but also enables the creation of more complex structures like protocol bridges. Also, the theoretical underlying model for IoTCapability mapping (typically

IoTSensorCapability to IoTControlCapability) was defined and explained with examples mapping different IoT protocols.

The work related to Component 5 ('Device Management and Orchestration') was not yet integrated in the pilot setup of the cycle 1 evaluation. Instead, it was conducted within the COMmeto laboratory environment and is planned for transfer to the pilot setup during cycle 2.

3.3.3 Pilot technical KPI measurements

Table 66 shows the technical KPIs of each component. For each of the KPI, validation metrics were defined using 3 levels. The goal is to reach level 3 at the end of the project. A distinction was made between KPIs targeted for cycle 1 and 2. Table 66 lists the relevant KPIs including the validation metrics that were targeted for cycle 1. A full overview of the validation metrics including those of cycle 2 can be found in D5.1, but are not repeated below. This section describes the results of each component during the pilot evaluation and discusses whether the targeted validation metrics for cycle 1 were met.

Regarding the sensor node design, each component used different development kits for communication, localization, or sensing. Processing was performed for each component separately. Cycle 2 will focus on running the proposed algorithms on embedded hardware and integrating the sensor node into a single design.

Although device-free person tracking was originally foreseen for cycle 1 (zone selection in a patient room at a range of up-to 5 meters with 30 cm distance accuracy), this was not yet investigated as the focus lied on activity classification models for both radar technologies. As a result, this validation metric will be transferred to cycle 2.

Table 6: Technical KPIs for P1-G.

Component	KPI	Validation metrics cycle 1
Design of next generation care communication platform based on **new sensor nodes that communicate wirelessly** to different network topologies and perform communication, localization and sensing functions	**Design of a sensor node** comprising a radar sensing module and a communication module that can run an embedded protocol stack	Hardware design using different development kits and a raspberry pi for processing.
	A joint communication and sensing network stack that shows how communication/ localization /sensing can optimally support various data exchanges dealing with varying application requirements and different network topologies	1) Provisioning protocol for devices, including discovery of capabilities and configuration or derivation of context. 2) Network orchestration to organize RF communication with consideration of application requirements and topology.
Hybrid indoor localization solutions based on BLE HADM (High Accuracy Distance Measurement), UWB and/or FMCW radar and understanding of the	**Indoor localization solutions** based on BLE HADM (High Accuracy Distance Measurement)	1) Ranging based on BLE HADM with accuracy of 15 cm in LOS and 30 cm in NLOS conditions. 2) Localization based on BLE HADM with multiple anchors deployed in realistic positions in case of a static node with an accuracy of 30 cm.

trade-offs in performance/deployment and cost	**Device-free person tracking** using UWB/FMCW radar, capable of tracking a person in a room	Zones selection in a patient room at a range of up-to 5 meters with 30cm distance accuracy.
Radar-based patient monitoring using FMCW/UWB radars and activity detection algorithms that are robust wrt changing environmental conditions. Data will be extracted from different sources going from radar sensors, indoor localization. The results can be interpreted by caregivers in decision-making and follow up how the patient/elderly is behaving and evolving.	**Design FMCW based human activity recognition algorithms with improved generalization performance** wrt changing environmental conditions where a) there is variability in the relative positions between radar and person, walls and furniture, and human anatomy and/or b) there are other concurrently moving objects.	Models should be able to classify human movement activities) with 95% accuracy under realistic conditions where: the room and person characteristics are different from those during training.
Develop multi-sensor processing tools and modular building blocks for diverse sensor fusion issues	**Robustness of fusion algorithm with respect to sensor failing or outages.** On sensor failure or outage, the fusion algorithm falls back on the available sensor, maintaining sufficient accuracy (graceful degradation).	1) Algorithm continues normal working on inducing single sensor outage (no received values). 2) Algorithm detects and manages inducing single sensor failure (unrealistic values, extreme noise, ...).
	Accuracy of fusion algorithm compared to output of individual algorithms.	Sensor fusion algorithm is implemented and finetuned.

3.3.3.1 Joint Communication And Sensing (JCAS) stack

During the pilot evaluation, the accuracy of location updates and personal alarms was measured. Based on the scenario description of activities, possible location updates include: hallway, care room, and bathroom. During each scenario, participants pressed the button on the tag to trigger an alarm. Table 77 shows the results of location update and personal alarm accuracy. Received and missed messages were logged, in addition to any spurious messages.

Results show that alarms are received in all but one scenario, yielding an accuracy of 97.22%. However, location updates show an overall accuracy of 66.30%. This might be related to several factors: US synchronization between beacons, missed BLE packets, and localization accuracy of the US system. Location updates in the care and bathroom show an accuracy of around 70%, with 7 spurious location updates to the bathroom, of which 5 were corrected later by an additional (correct) location update. This can be explained by the fact that the two beacons in the care and bathroom are located close to each other. Due to a hysteresis mechanism in the US localization, this may yield incorrect and spurious location updates.

Locations in the hallway show a significantly lower accuracy of 47.83%. As the BLE gateway is located in the hallway, missed BLE packets are unlikely. However, both the BLE gateway and the

hallway beacon transmit US signals, which might suspect synchronization issues. In cycle 2, both the synchronization and BLE communication will be improved to target an increased accuracy.

Table 7: Accuracy of location updates and personal alarms.

Event	Received	Missed	Accuracy	Spurious	Corrected spurious
Care room location update	34	12	73.91%	0	0
Bathroom location update	16	7	69.57%	7	5
Hallway location update	11	12	47.83%	0	0
Overall location update	61	31	66.30%	0	0
Personal alarms	35	1	97.22%	0	0

As discussed above, the distribution of sensor data was not yet taken into account in the pilot evaluation, because the radar data was processed offline. However, D2.7 featured an evaluation of the available BLE throughput under the application requirements of sub-second alarms and ms-level synchronization between beacons. Under realistic interference conditions, a BLE throughput of 200-300 kbps was found achievable.

To target the provisioning protocol of devices, several commissioning protocols were considered, with or without a commissioning agent and whether or not BLE localization was part of the commissioning protocol. Two initiating procedures were developed for BLE localization, to discover nearby anchor nodes and schedule ranging interactions without collisions. The commissioning protocols and initiating procedures were developed theoretically but were not yet part of the pilot evaluation, as the evaluation focused on the operational state of the JCAS, not including discovery and configuration.

3.3.3.2 BLE localization using channel sounding

Ranging was performed throughout the entire room. However, the accuracy of both ranging and localization is highly dependent on the relative position of the tag and the anchor and its orientation. Localization errors increase significantly when ranging accuracy decreases, as localization is directly dependent on ranging and errors are amplified through trilateration. Therefore, the primary focus of this validation cycle is on assessing ranging performance in a complex indoor environment without controlling the device orientation.

In Cycle 2, additional anchors will be introduced to evaluate the impact of anchor quantity and placement on overall system accuracy.

Figure 12 presents the cumulative distribution function of ranging errors for all measurement points with known ground truth. Approximately 70% of the ranging measurements, LOS and NLOS, achieve an accuracy of under 1 meter, which is acceptable for many indoor applications, but not precise

localization. Additionally, at least 20% of the measurements, of which 60% are NLOS, exhibit errors of 2 meters or more, which severely affect trilateration. These large errors can result in completely incorrect position estimates, sometimes placing the tag on the opposite side of the room, several meters away from its true location.

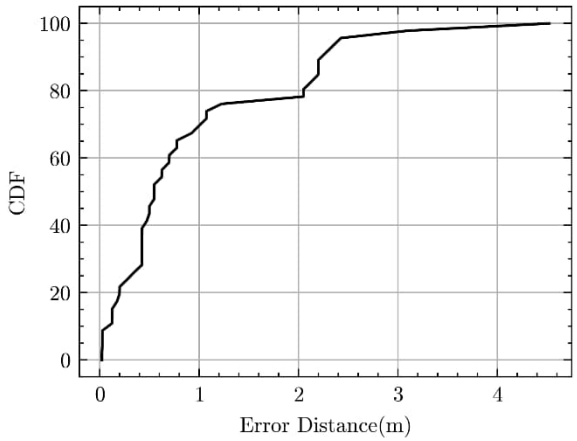


Figure 12 Combined LOS and NLOS distance error for all validation scenarios

Additional factors contributing to localization inaccuracies include the polling rate of ranging measurements. Internal errors occasionally lead to corrupted or missing data, causing certain anchors to update less frequently. This inconsistency further degrades the accuracy of the system.

In the next cycle, more advanced algorithms, such as Kalman or Gaussian models, will be explored to dynamically handle such errors. A more robust scheduling system will also be implemented to improve the reliability of Bluetooth connections used for PBR. These challenges underscore that the system's performance is still highly variable and sensitive to algorithmic choices and device orientation.

3.3.3.3 UWB radar activity detection and tracking

This analysis examines sleep-related activities using UWB radar. We define four labels: No activity, Anxious bed movement, Sleep interruption event, and Nocturnal wandering. Anxious bed movement is when the person is moving while lying in bed, a sleep interruption event corresponds to the person standing up from or sitting down on the bed. The ceiling-mounted radar best captures movements in and around the bed, but reflections weaken with distance, making far activities harder to detect. This explains the focus on near-bed activities. To capture finer room-level activities in cycle 2, four additional receivers will be installed in the corners of the room. In future data campaigns, these will record CIRs from a more optimal angle.

Figure 13 shows a confusion matrix of currently achieved accuracies. The model predicts "No activity" with 100% precision, yet many true "No activity" samples are misclassified as a "Sleep interruption event". Because interruption events are short and fast, they are difficult to label precisely, and some actual "No activity" samples may have been mislabelled. In contrast, "Anxious bed movement" involves continuous motion and is easier to label, which is clearly shown in the matrix. "Nocturnal wandering" is also challenging, as pauses or movements outside the radar's view can resemble inactivity. Correlation-based algorithms have improved labelling accuracy, but further refinement is needed.

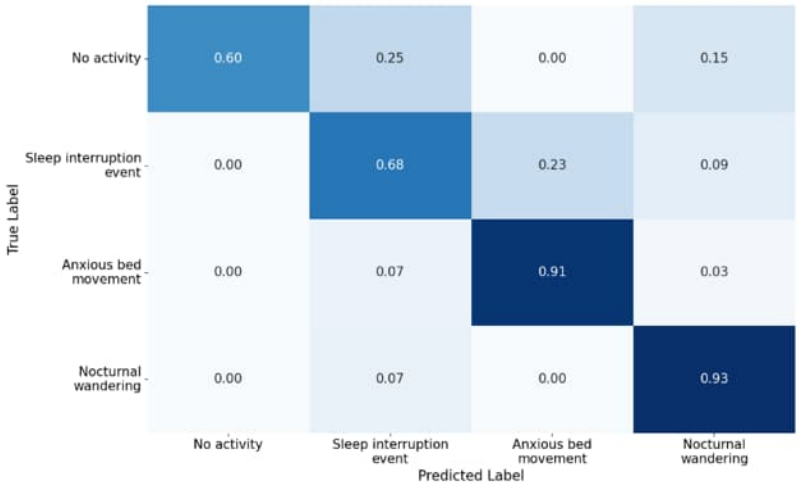


Figure 13 UWB confusion matrix for all validation scenarios

Validation on scenarios 1 and 2, which most closely match the training data, yields an F1 score of 93%. Scenarios 5 and 6 achieve 85%, while scenarios 3 and 4 perform significantly worse at 56%. These differences highlight the need for better generalization, either through more location-independent preprocessing or greater scenario diversity during training. Important to note is that scenarios 4 and 5 had the greatest distance between person and radar, possibly approaching the maximum effective range of the system.

3.3.3.4 FMCW radar activity detection and tracking

We evaluate human activity recognition algorithms using FMCW radar technology, with a specific focus on assessing the generalizability of deep learning models across different scenarios and environmental conditions. The primary objective is to demonstrate the robustness of a convolutional neural network approach when deploying models to handle domain shifts in activity recognition scenarios. The first evaluation encompasses nine distinct human activities: walking, sitting down, lying down, standing up, getting up, eating hand movements, drinking hand movements, turning in bed, and anxious movements on the bed. The second evaluation that is performed on the validation set focuses on a reduced set of seven activities: walking, sitting down, lying down, standing up, getting up, eating movements, and anxious movements. We employ a CNN-based classification model featuring four convolutional layers with batch normalization, ELU activations, and dropout regularization, followed by two fully connected layers for feature extraction and classification.

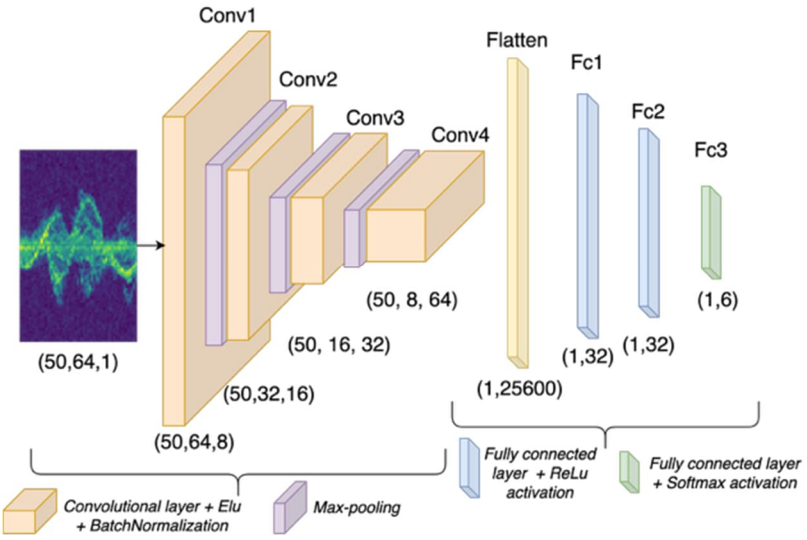


Figure 14: CNN model architecture

Data collection was carried out within the IDLAB environment using two distinct evaluation methodologies. The first evaluation employs a Leave-One-Subject-Out (LOSO) cross-validation approach, where models are trained on data from all participants except one, which is reserved for testing. This process is repeated for each participant, and the results are aggregated by summing all confusion matrices to provide comprehensive performance metrics. The second evaluation method involves training models on data collected during the initial training phase and validating them on a newly collected validation dataset. This validation dataset (Figure 14) was specifically gathered using a different set of scenarios while maintaining similar activities and features different room layouts compared to the training environment, thereby testing the model's ability to generalize across spatial configurations and scenario variations within the same facility.

The evaluation is conducted using three key metrics: F1-score, accuracy, calculated as a weighted average across all the classes, the confusion matrix, which provides detailed insights into per-class classification performance.

Leave-One-Subject-Out evaluation

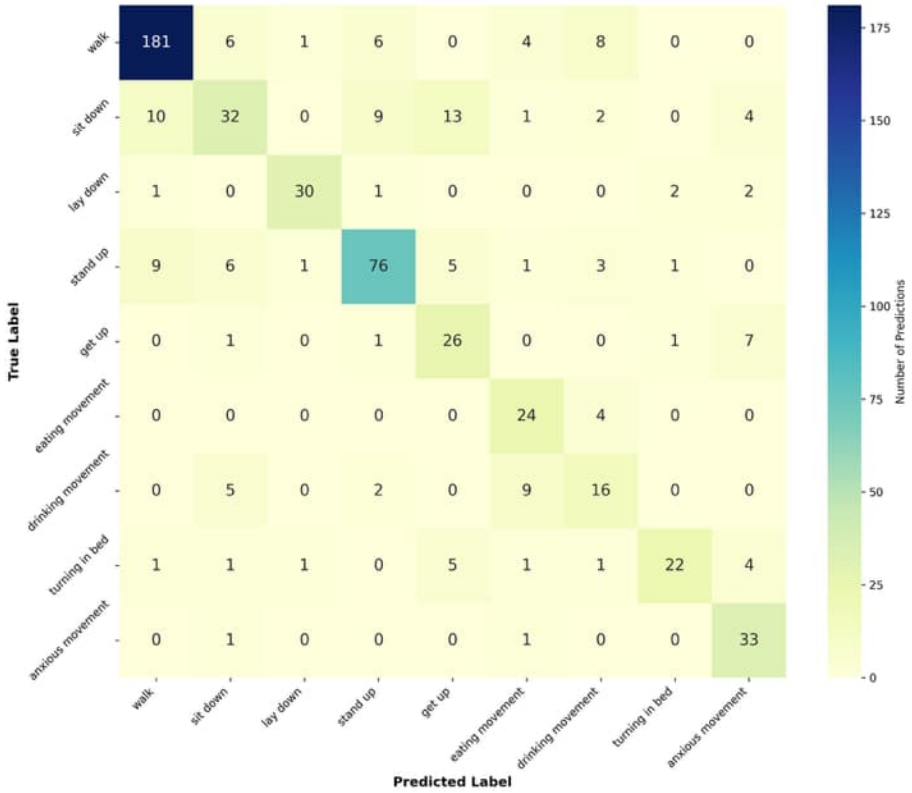


Figure 15:Confusion matrix for Leave-One-Subject-Out validation

The CNN model achieves an overall accuracy of **76.4%**, with a macro-averaged F1-score of **0.7030** and recall of **0.7268** across the nine human activities. The confusion matrix reveals predictable classification challenges where activities with similar movement patterns generate overlapping time-doppler signatures. Most notably, eating and drinking hand movements show significant confusion due to their similar hand dynamics, while postural transitions (sit down, stand up, get up) exhibit cross-classification errors reflecting shared biomechanical initiation phases.

Evaluation on newly collected validation set

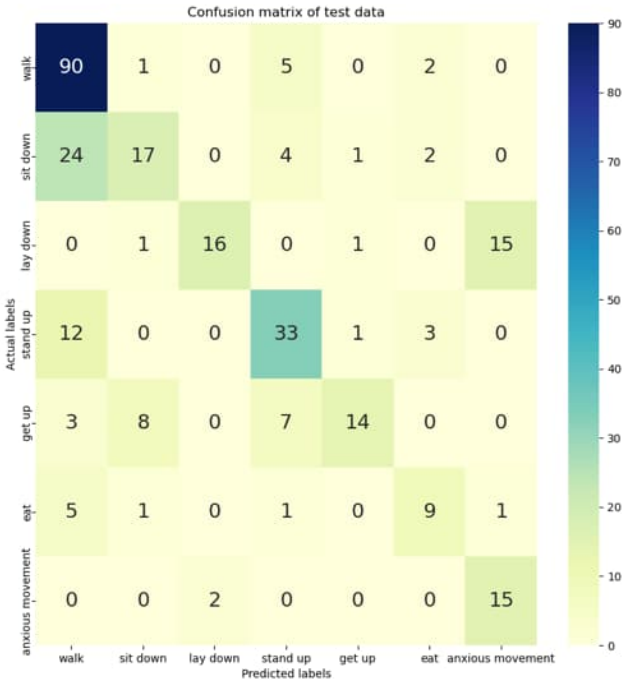


Figure 16: Confusion matrix for validation dataset

The Figure 16 displays the confusion matrix results from newly collected validation data. This dataset incorporates novel room configurations and includes four participants who were not present during the model's training phase. These factors contribute to a notable performance decline between the first and second experiments, with the F1-score dropping to 54% and accuracy falling to 63%. Such degradation is expected, as time Doppler spectrograms are sensitive to an individual's position during activity execution, given that they capture only radial velocity measurements. Additionally, the validation dataset contained participants with demographic characteristics different from those in the training set, which further impairs model performance. Nevertheless, this challenge can be addressed through data augmentation techniques specifically designed for FMCW radar data, combined with training approaches such as unsupervised domain adaptation.

3.3.3.5 Multi-sensor fusion

We have made progress primarily on early fusion techniques, as we believe this is the most interesting direction to go in. However, as of yet, we are still working on the implementation. Hence, we cannot provide our defined KPI’s in this cycle.

3.3.3.6 Device management

COMmeto has made significant progress toward the KPI on "Applicability of device management on the edge devices in the distributed platform", focusing on the "Gateway setup to process the data produced by the sensors and send these data to the cloud." Our work in the first cycle created a crucial foundation for this objective by developing a new data model and conducting key lab experiments.

In our lab, we successfully conducted initial experiments that demonstrated the applicability and interoperability of this model. We achieved promising results with a variety of protocols, including LoRaWAN, Matter, Loxone, Zigbee, MQTT generic payload and AMQP generic payload. Furthermore, we successfully tested the applicability of the concepts applied in the interaction between USP-Controllers and UPS-Agents. We also formally defined a model for IoTCapability mapping (IoTSensorCapability to IoTControlCapability), which provides the foundation for seamlessly connecting different IoT protocols.

In the second cycle, COMmeto will move from lab validation to full pilot integration, ensuring that device management is fully embedded in the gateway. This will enable secure sensor-to-cloud data flows, remote device reconfiguration, and advanced interoperability across heterogeneous sensor networks.

3.3.4 User aspects: stakeholder engagement in pilot development

We involved both stakeholders and target users through a focus group discussion. The target users of this pilot are residents and staff (caregivers and management) in a care facility. The aim of the focus group was to capture expectations, understand needs, and gather reflections on the relevance and acceptability of lifestyle monitoring technology in daily care practices.

A total of nine participants contributed:

- 4 end users, representing residents and staff (caregivers and management).
- 5 technology partners, representing the technical development side of the project.

Among the end-user representatives, several stakeholders with key roles in the care facility were engaged, including care management, a head nurse, a therapist, and an IT system responsible. Their involvement ensured that perspectives from daily care operations, clinical practice, and digital infrastructure were all considered.

The focus group itself was organized by two technical partners, who introduced the DistriMuSe project, the pilot objectives, and the intended benefits for care delivery. The session began with a short presentation, followed by an interactive polling and feedback tool that enabled participants to provide input anonymously and visualize results in real time. This approach encouraged active participation and triggered open discussion.

End users were asked, among others:

- *“What are the greatest needs as a caregiver to provide optimal care?”*
- *“What comes to mind when you hear the concept of lifestyle monitoring?”*

To make the dialogue more concrete, two situational sketches were presented:

- Being the only caregiver on duty during a night shift.
- Drafting a care plan for a resident.

The answers and comments were used to stimulate debate and to better understand where lifestyle monitoring could provide the most value. For the remaining technical partners, input was collected separately using a questionnaire. Their responses were used to complement the insights gathered in the focus group and to ensure that ethical and design considerations were also reflected from a technological perspective.

3.3.5 User-based KPI assessment

The relevant KIPs from an end user perspective were defined in D5.1. For each of the KPI, validation metrics were defined using 3 levels. The goal is to reach level 3 at the end of the project. A distinction was made between KPIs targeted for cycle 1 and 2. Table 8 lists the relevant KPIs including the validation metrics that were targeted for cycle 1.

Table 8: User-based KPIs for P1-G.

KPI	Validation metrics cycle 1
The system shall be able to classify multiple activities under realistic circumstances. When multiple persons are present in the room the algorithm should not make predictions.	1) Activity classification for single person: coarse grained
The system shall be able to classify multiple activities under realistic circumstances in different rooms	1) Activity classification for single environment 2) Activity classification for a second environment
System is validated by end users	1) Validated based on data collected by researcher, evaluated technically 2) Validated based on data from realistic persons in a living lab, evaluated by end users
Number of stakeholders included	1) Professional caregiver

According to the technical evaluation of the different components in Section 3.3.3, coarse grained activity detection for a single person is feasible, albeit with varying accuracy depending on the type of activity and distance from the radar. Nonetheless, multiple activities can be classified in a single environment, with reduced accuracy in a second environment that features other room layouts, a different activity sequence, and new test persons. The system was validated based on data from researchers and realistic persons in two data campaigns, but not yet evaluated by end users. Several professional caregivers, including a head nurse and therapist, were involved in a focus group discussion.

3.3.6 User aspects: Gender/age issues and ethical concerns in P1-G

As described above, the ethics focus group has been completed. The session was conducted with a total of nine participants, including target users (care staff and management) and technical partners. Care facility representatives were included to ensure that practical, clinical, and organizational perspectives were reflected, while technical partners were involved to capture the development-side considerations and responsibilities.

During the session, end users and care stakeholders participated in a guided discussion facilitated by the ethics exercise, while technical partners who could not attend the meeting provided their input separately through the structured questionnaire. This approach ensured broad representation of both user needs and technological viewpoints.

The outcomes of the exercise will be included in D7.7. The focus group has already provided useful input for reflecting on our pilot's evaluation strategy and highlighted aspects to be considered in the next phases of implementation.

3.3.7 P1-G transition into cycle 2: takeaways and feedback

Cycle 1 validated individual sensing and communication components in a realistic living lab, including FMCW and UWB radar for activity recognition, BLE channel sounding for localization, JCAS protocols for reliable alarms, sensor fusion concepts, and device management extensions. The components demonstrated feasibility, while stakeholder focus groups confirmed the relevance of lifestyle monitoring technologies.

However, radar models showed limited generalization across new rooms and participants, and BLE localization produced large errors in NLOS conditions. JCAS location updates lacked reliability, and sensor fusion was not yet validated. Device-free localization using UWB or FMCW has not yet been explored, and device management was not yet integrated into the pilot setup.

In cycle 2, all components of the room sensor node will be integrated on embedded hardware. BLE HADM localization will be improved by fusing complementary techniques such as device-free localization (using either UWB or FMCW), point-cloud–based methods, and ultrasound localization. JCAS will be extended with commissioning and scheduling for BLE HADM, while also improving reliability. Fusion algorithms will be implemented and tested in realistic pilot conditions, including investigations into moving from coarse-grained activities toward higher-level activities (e.g., combining location information with coarse actions to infer "eating at the table"). More extensive data campaigns will be carried out across multiple setups while capturing a wider range of activities to improve robustness and generalization. Device management will be deployed in the pilot environment, and care staff and residents will directly test the system in real scenarios, guiding specifications for next-generation, scalable communication and sensing nodes.

3.4 P1-KMPHG evaluation cycle 1

Sleep is an important factor for the common well-being and health. In general, one in ten people have severe problems with their sleep, such that it influences their daily functioning. Video-polysomnography measured at a sleep laboratory is the gold standard for the objective evaluation of sleep and the diagnosis of sleep disorders. This is a rather obtrusive method, using many wired sensors, and is mostly performed in specialized clinical settings.

The sleep pilots are run at two study sites:

- P1-KPMHG: Patients suspected to have a sleep disorder, within the Centre for sleep Medicine Kempenhaeghe, Netherlands).
- P1-KSL: Healthy volunteers without diagnosed or self-reported sleep disorders, SmartSleep Laboratory at Kuopio University of Eastern Finland

The objective is to develop less obtrusive sleep monitoring technologies that offer reliable results, serving as alternatives to the current gold standard — video-polysomnography — by using less invasive methods. Furthermore, this might result in more reliable results, as the patient does not feel monitored or restrained during the night, and therefore the measured data might be closer to reality.

3.4.1 Final pilot set-up

P1-KMPHG is performed within the Centre for sleep Medicine Kempenhaeghe, a tertiary referral centre specifically for patients with severe or complicated sleep disorders. Assessment of sleep is done using a video-polysomnography, the gold standard method for measuring sleep. Each polysomnography takes place in-hospital. Patients arrive in the late afternoon, and the measurements setup will be prepared.

The proof of concepts at P1-KMPHG include:

- eLive – Edge system
  - Connecting, data processing, recording and control of the sensors such as the Bed sensor and VTT radar
  - Launching and monitoring application at the control room
  - Provides external synchronization signal for the reference PSG recordings

- eLive – Bed sensor
  - A multichannel foil sensor placed in below the bed mattress
  - The system processes as on-line the vital signs and sleep features and stores both the results and the raw sensor data
  - Connected with the eLive-Edge system

- VTT – Radar
  - Frequency Modulated multichannel radar at 60 GHz range placed high at wall targeting at the whole area of the sleeping subject at bed
  - The online processing runs the FFT-range 2D-image formation and stores the ROI area of the sleeping subject for the further off-line analysis of vital signs and sleep.
  - Connected with the eLive-Edge system

- Infineon – Radar
  - Continuous-wave 60 GHz radar placed at ceiling above the subject targeted at the middle body
  - Started and stopped from a Raspberry Pi with touchscreen. After each night a backup is made to a USB device. The radar is running in CW-mode and mounted to the ceiling.
  - Provides external synchronization signal for the reference PSG recordings

The KMPHG sleep pilot architecture is in the next Figure.

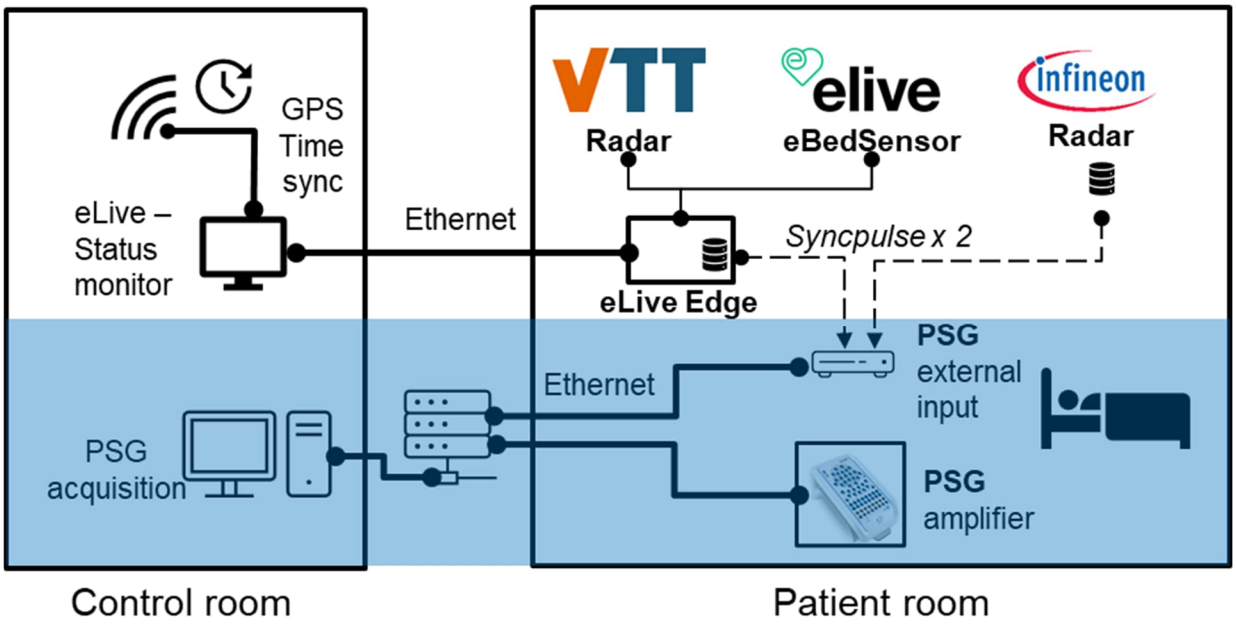


Figure 17 P1-KMPHG sleep pilot architecture. The DistriMuSe study demonstrator is at top, and the hospital PSG-system below is covered with dim blue

The study devices at the patient room on the right-hand side of the figure, include the VTT radar, eLive eBedSensor and Infineon Radar. The pilot system status monitor is in the sleep laboratory Control room on the left-hand side.

The VTT radar and eBedSensor are connected to the eLive Edge device. The recording for these is operated using the status monitor at the control room with Ethernet connection to the eLive Edge device. The system clock can be synchronized with GPS-USB module.

The recorded data for each sensor is stored locally and transferred with external USB-connected SSD disc. Sensor data can be re-synchronized accurately with the PSG recording with the external syncpulse from each sensor system, which are recorded from the PSG external input.

Next Figure shows the VTT and Infineon radar placement at ceiling above the bed.

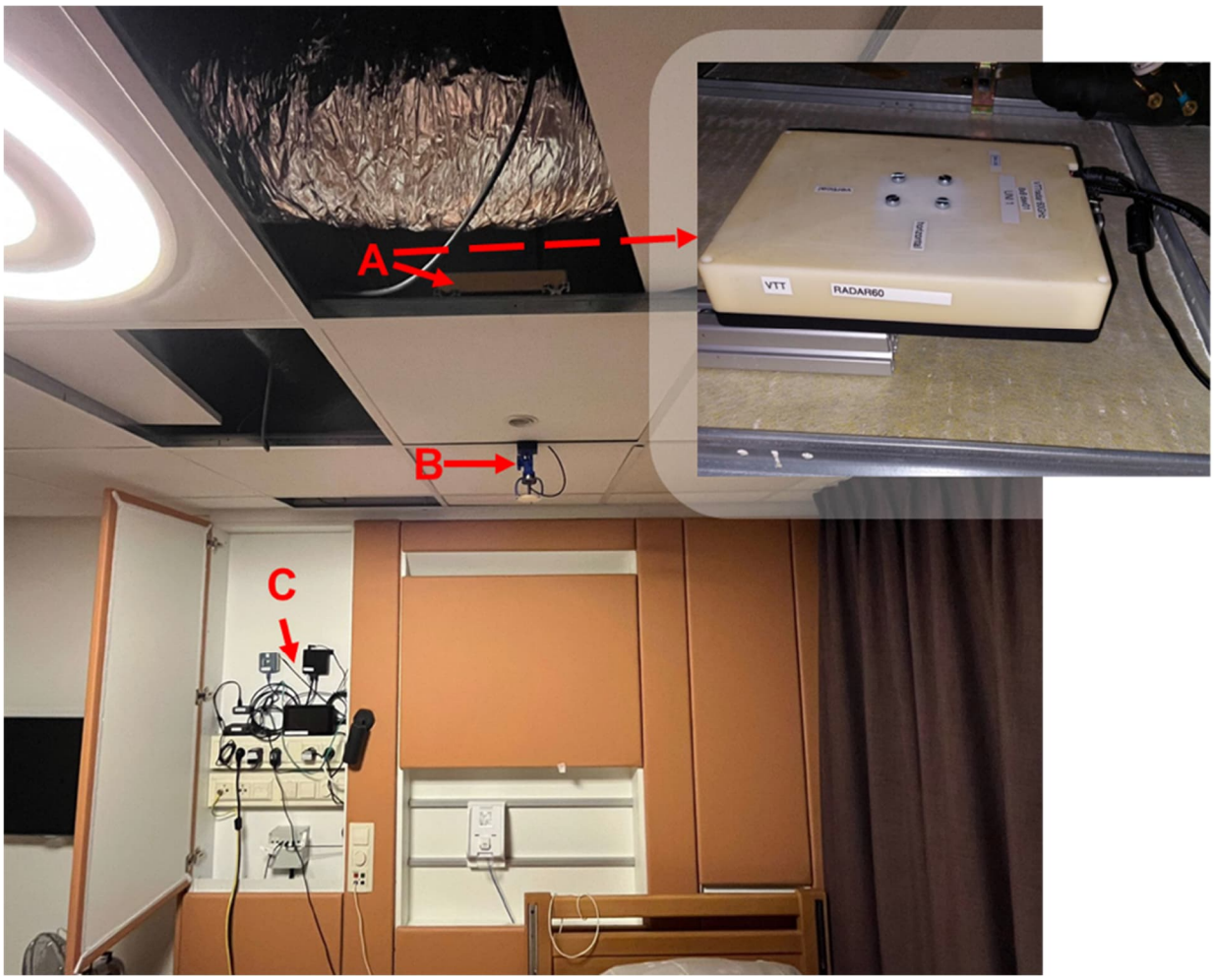


Figure 18 VTT radar is placed above the ceiling plate and marked with 'A' in the picture. The VTT radar from above is illustrated with the added picture in the right-hand side. The Infineon radar is fixed to the ceiling metallic grid rail and marked with 'B'. The eLive Edge and Infineon control devices are placed inside the cabinet as marked with 'C'

The Infineon radar at the ceiling is shown in the next Figure when the ceiling plates are in place.

Figure 19 When the ceiling plates are in place, the VTT radar is hidden above, and the Infineon radar is as shown in the picture.

Measuring through the hollow ceiling plates does not affect the VTT radar measurement sensitivity.

The eLive eBedSensor is placed in below the bed mattress as shown in the next Figure.

Figure 20 eBedSensor below the bed mattress.

3.4.2 Pilot evaluation execution and protocol details

3.4.2.1 eLive EDGE system

Each demonstration will leverage the eLive EDGE platform to synchronize and preprocess under-mattress and radar data streams, performing multi-sensor fusion in near real time. This ensures that the combined outputs - respiration rate & flow, heart rate, posture, and leg-movement flags - can be

benchmarked against the gold-standard PSG. Success will be assessed against the performance thresholds (accuracy, latency, and robustness).

The system components are connected with Ethernet via the LAN router. The edge device RPI5 (RasperryPi 5 unit) is connected with the sensors and handles the recording and preprocessing of the data. The edge device also generates the synchronization signal, which is connected as an external signal for the sleep laboratory PSG data recording, to enable accurate data synchronization for the offline analysis between the sensors and the reference PSG.

The user interface for the sleep technicians is provided by an additional Raspberry Pi unit at the control room. The main screen is shown in the next Figure.

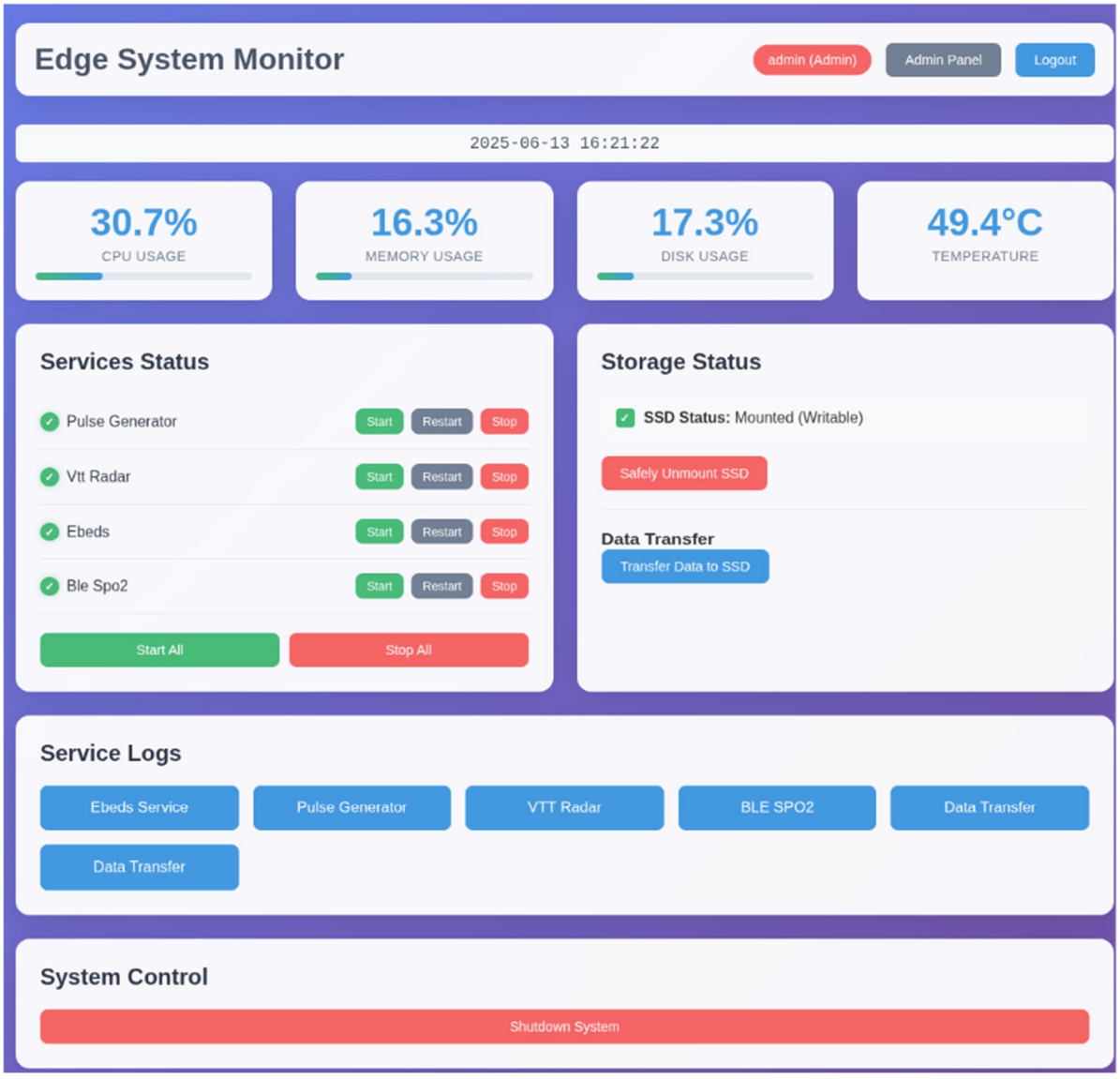


Figure 21 eLive-Edge operation GUI

Top part shows information for the edge RPI5 device such as the device usage and CPU temperature. The left side has control buttons for each sensor and the pulse generator (DAQ). Right side has button for data transfer to the removable SDD drive. In below the user can open more detailed service logs showing performance for each system component, e.g. the number of missed

samples during the current overnight recording period. The lowermost button runs the safely shutdown for the system before powering off.

3.4.2.2 eLive eBedSensor

eBedSensor is under-mattress sleep monitor with multiple measurement points for vital-sign monitoring. It is including:

Respiration monitoring: Validate continuous, unobtrusive measurement of respiration rate and flow beneath the mattress, comparing against PSG airflow channels and respiration belt signals.

Sensor fusion support: Provide data from our under-mattress sensor to enhance radar analytics over radar-only measurement by delivering precise timing of respiratory and movement events, thereby streamlining the development of multi-sensor fusion algorithms. Investigate how sensor fusion improves performance compared to individual measurements alone.

Offline operation & data off-load: Prove that the sensor + eLive EDGE platform can record a full night (> 8 h) without network connectivity and then upload all data within 10 minutes at < 1 MB/min average bandwidth.

3.4.2.3 VTT 60 GHz FMCW Radar

VTT has developed 60GHz imaging radar for people motion tracking and remote sensing of the human vital signs. The image formation is with 2D, first for the distance in meters and then for the horizontal angle. The radar is using Frequency Modulated Continuous Wave (FMCW) principle with eight transmit and eight receiving antennas, and the multiple input/output (MIMO) processing enables simultaneous sampling within each image frame.

For the second project cycle the radar shall be expanded to the fully polarized system. This shall provide more information about the target micromotion and more robust sensing in the multipath indoors environment.

3.4.2.4 Infineon Radar for vital sign monitoring

Infineon has developed a 60GHz radar device in continuous-wave mode to detect and extract heartbeat signals from raw data. The data will be recorded throughout the entire night and synchronized with the reference PSG system.

3.4.3 Pilot technical KPI measurements

The relevant project KPI for both the P1-KMPHG and P1-KSL sleep demonstrations are listed in the next Table.

Table 9 Technical KPIs in P1-KMPHG

KPI Topic	Innovation	Validation status
Radar-based sensing for physiological signal monitoring	We will improve the sensitivity and robustness of the radar. In addition, we will study what kind of the new possibilities radar polarimetry could bring to the table in health applications in general.	Radar HR and respiration signals are compared against the PSG references with the first recordings from both P1-KMPHG and P1-KSL pilots. The radar polarimetry shall be studied in the project cycle 2.
Imaging radar-based sensing for human motion detection	We intend to improve the resolution of imaging radars as well as experiment with a novel polarization detection approach, to better identify body part movements e.g., in sleep monitoring.	Same as above.
Early in-sensor fusion	Each remote sensor is measuring the micromotion of the skin surface from different directions and at different locations of the body, and so forth the sensor fusion can improve sensitivity and robustness especially for the heart rate and the pulse arrival time (PAT) measurement.	The radar, eBedSensor and the pulse oximetry signals are combined with the accurate time synchronization and the first estimations of the PAT is reported in her the P1-KSL pilot section.
Late fusion	Sleep modelling and sleep disturbances detection based on multimodal sensor information from radar, bed mattress sensor and wearable polygraphy system	Sleep modelling is not yet statistically quantified because of the small number of recorded nights.
Distributed computing	Radar and sensor signal processing in edge devices and data fusion in the common platform. Data synchronization.	The VTT radar and eBedSensor are recorded and on-line processed into physilogical signals with the eLive Edge device. Time difference with the PSG system is fixed based on the external synchronization signal.

This section shows results corresponding the heartrate measurement and the time offset synchronization.

Next Figure shows the heart rate measurement with the VTT radar and eBedSensor in reference with the ECG based measurement of the PSG system.

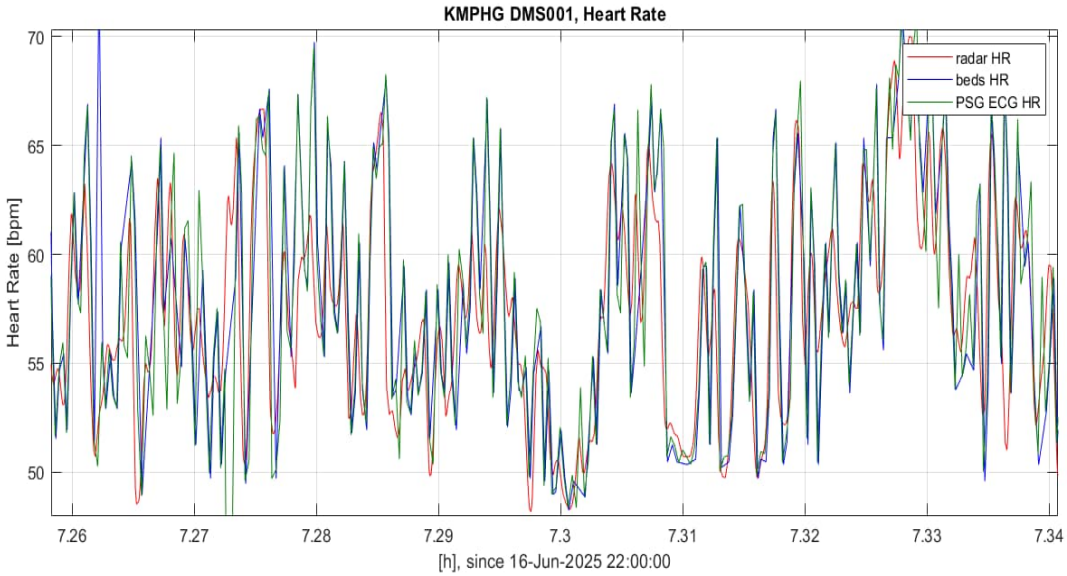


Figure 22 Heart rate measurements from the radar (red signal) and eBedSensor (blue signal) match well with the reference ECG (green signal). About 10 minutes period during sleep is shown for the subject id DMS001 at KMPHG sleep laboratory.

The VTT radar and eBedSensor recording timestamps are in-sync because they both measured with the eLive Edge device and labelled with the same system clock.

The synchronization pulse signals from the eLive Edge and the Infineon Radar enable accurate offline synchronization with the reference PSG signals. Next Figure shows the found time offset difference between each system.

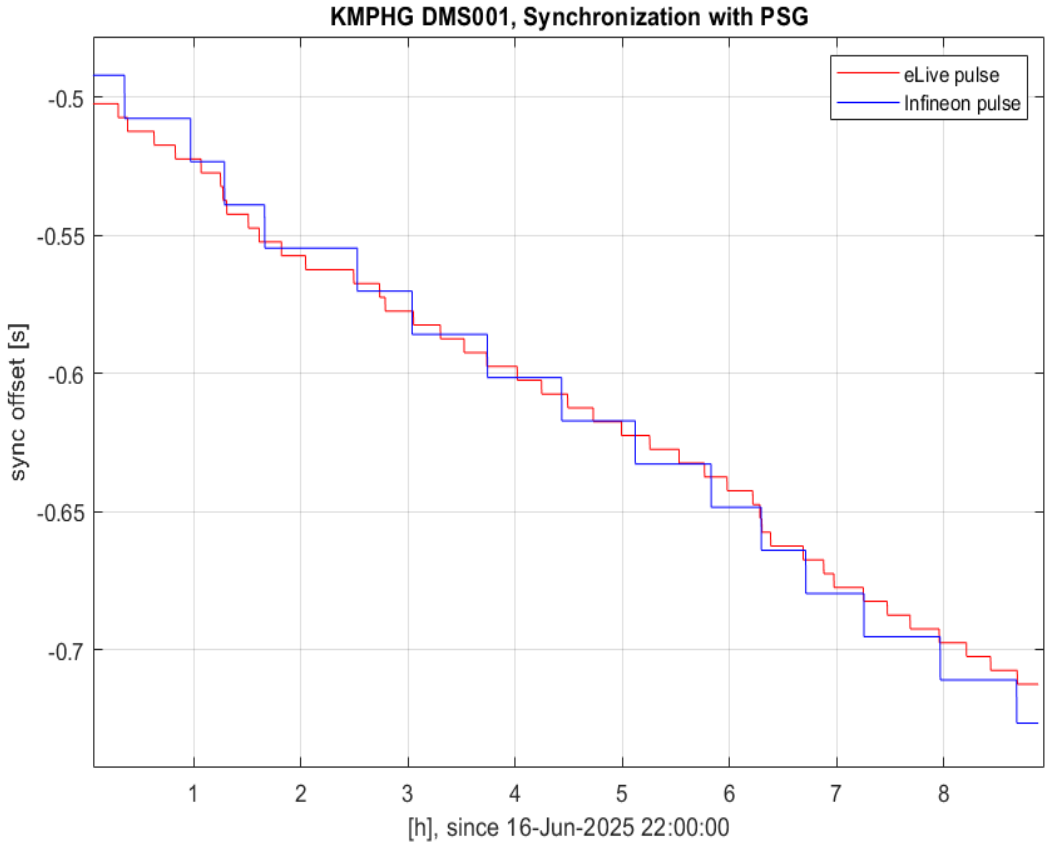


Figure 23 Synchronization time offset between the eLive and the Infineon systems against the reference PSG recording.

For the example in the Figure, the offset difference seems to drift from –0.5 seconds down to –0.72 seconds during the nine hours of the overnight sleep recording. The offset is measured by cross-correlation between the pulse signal from the corresponding recordings. Result is quantized with about 10 milliseconds resolution for the eLive system and about 30 milliseconds resolution for the Infineon system. However, as the offset is seemingly drifting slowly, the resolution can be improved by smoothing and interpolation methods. In this case the eLive and Infineon offset seems to drift in similar manner against the PSG. Both eLive and Infineon are RaspBerry 5 devices and as they are in the same cabinet the temperature effect is similar for each system clocks.

Minimal offset difference is important especially for the heart Pulse Arrival Time calculation (PAT). The heartbeat detection and thus the PAT can be estimated from the VTT radar and eBedSensor with two different principles:

- FFT-based sliding window Cepstrum method
- Time based triggering on each heart pulse

The latter approach seems to be more similar with the reference method for detecting the ECG-R peaks. However, the shape of the heart pulse variates strongly e.g. relating to the respiration phase, i.e. inhaling or exhaling, and the triggering on the heart pulse does not provide as accurate heartbeat interval estimate as the FFT-based method. The FFT-based method performs the channel wise

averaging on the Power Spectra in the frequency domain and thus the phase variance of the heart pulse signals does not affect the pulse detection. (Bruser et.al. 2015)

Next Figure shows the PAT estimation for the participant DMS001.

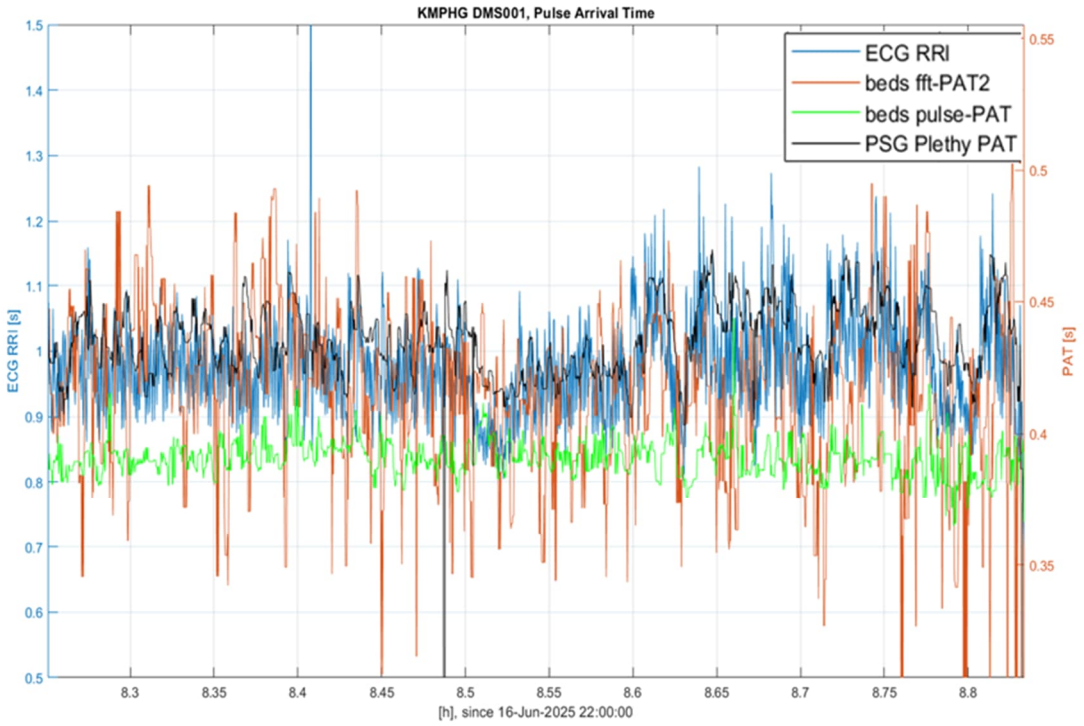


Figure 24 Left side axis shows the ECG R-R interval in seconds with the blue curve. The right-side axis shows two different PAT estimates from eBedSensor (red and green curves) and from the reference PSG photoplethysmography (black curve).

The reference ECG R-R interval (ECG RRI) is shown firstly in the comparison above, because it is well known that the PAT normally correlates with it. The black colored reference PAT measured between the ECG and the PSG's photoplethysmography signal (PPG) seems to follow the blue coloured RRI curve as it should. While the ECG RRI range in the left-hand side axis is about 900ms ... 1100ms, the PAT range in the right-hand side is 400ms ... 450ms. Typically, the change of RRI is somewhat larger in seconds than the corresponding change in PAT. (Peng et.al. 2021)

In the Figure, the FFT-based PAT with red color seems to follow the reference (PSG plethy PAT with black color) better than the heart pulse detection-based PAT with green color. So forth, the FFT-based pulse detection approach is suggested both for the heart rate and the PAT estimation. This shall be proved by end of the cycle 1 sleep pilots at P1-KMPHG and P1-KSL.

3.4.4 User aspects: stakeholder engagement in pilot development

As stakeholder engagement is critical in the development of innovative solutions. During this pilot, stakeholders and users where those who are running the measurements in the clinical setting. This includes the sleep lab technicians as well as the lab manager and coordinator research. As well as patients and/or subjects undergoing measurements with the newly developed proof of concepts.

To gain insight into stakeholder perspectives and gather feedback on the proof of concepts, any issues with the different proof of concepts were noted by the sleep lab technicians and when possible, linked to one of the several proof of concepts.

During the pilot, none of the patients who underwent the simultaneous recording of both the clinical polysomnography and the measurements with the different proof of concepts, mentioned or complained about the addition of the different proof of concepts. The lab technician and lab manager encountered several issues while using the different proof of concepts, by direct calls and local visits of the technical team members to the pilot site technical issues were solved. As well as smaller updates to the measurement protocol and software were done, to meet the clinical workflow on the pilot site.

3.4.5 User-based KPI assessment

Table 10 User-based KPIs in P1-KMPHG

KPI	Validation metrics cycle 1
System should not disturb sleeping by any manner	The overnight measurements done so far in two sleep laboratory pilots did not result any harm for the test participants. Especially the non-contact radar and eBedSensors are fully unobtrusive. The wearable Bittium respiro system at P1-KSL pilot is additional for the PSG sensors and thus might add some discomfort, e.g. as having oximeters in multiple fingers.
Accurate time synchronization is needed for the sensor-signals level fusion, synch is not so critical for the sleep parameters level fusion	The eLiveEdge system in development shall provide accurate synchronization, as it is off-line tested for the current recordings.
Physiological parameters processing online at edge to reduce data transfer and storage	The radar and eBedSensor systems data preprocessing is performed online at the present eLive Edge system. This includes the radar image formation and the physiological feature extraction for both sensors.
For the sleep reporting the processing can be done off-line right after the wake up	The eLiveEdge system performs currently the eBedSensor sleep analysis and reporting automatically after the sleep session.
Specific applications may benefit from on-line sleep modelling, e.g. sleep optimization, apnea prevention	The sleep analysis is currently run offline after the sleep sessions to enable more accurate filtering and calibration procedures with non-causal approaches. However, some of these can be directly turned into on-line processing. A feasible latency time for processing is assumed to be within about ten seconds for apnea prevention and several minutes for the sleep optimization.

3.4.6 User aspects: Gender/age issues and ethical concerns in P1-KMPHG

As part of the preparations on the pilot, an ethics focus group meeting was conducted on the 4th of August 2025. During the focus group, 7 participants were present from the pilot site as well as from the different proof of concept developers. By design, the different proof of concepts have many advantages over the current clinical practice, any now serious drawback could be identified during the focus group. Data is stored in a completely secure manner; the collection of the data is less obtrusive and maybe more representative of patients' actual way of sleeping.

The data collection of the pilot P1-KMPHG was running from mid June 2025 until September 2025. All patients gave written informed consent before participating. In total 28 participants were included, of which 15 were female and 13 were male. The measurements meet all local standards and legislation. No protocol violations were reported. All data of the different proof of concepts was stored pseudonymized. Unfortunately, in at least 7 measurements, data was lost due to technical issues or user errors. Further analysis of the data is needed to ensure the quality of the data.

3.4.7 P1-KMPHG transition into cycle 2: takeaways and feedback

- In total 33 participants were recruited, in which both clinical polysomnography as well as the different proof of concepts were used to simultaneously record different vital sign parameters during sleep.
- User errors and technical errors caused data loss in at least 7 participants.
- Intended changes for cycle 2 focus on an even more robust measurement setup (less data losses), tweaked measurement setups based on the results of the first pilot. Additional sensing for the continuous measurement of blood pressure.

3.5 P1-KSL evaluation cycle 1

Because the sleep pilots P1-KMPHG and P1-KSL are closely related, only differences compared to P1-KMPHG will be mentioned in this section for the pilot P1-KSL.

3.5.1 Final pilot set-up

Measurements are performed in SmartSleep Laboratory at University of Eastern Finland, Kuopio. Data will be collected simultaneously from the proof-of-concept devices and gold standard PSG. The first tests of the P1-KSL protocol were conducted in May 2025. The pilot is planned to continue in September 2025. However, the actual start of the measurements is dependent on the Finnish Medicines Agency (FIMEA) approval for medical device study.

In contrast to P1-KMPHG, only healthy individuals are recruited for this study. In addition, the Finapres CNIBP and the Bittium Respiro HSAT devices will be utilized.

In P1-KSL the VTT radar is installed on the wall with a tilt angle (Figure below). The radar is fixed with an adjustable arm to enable FOV positioning to upper part of the body.

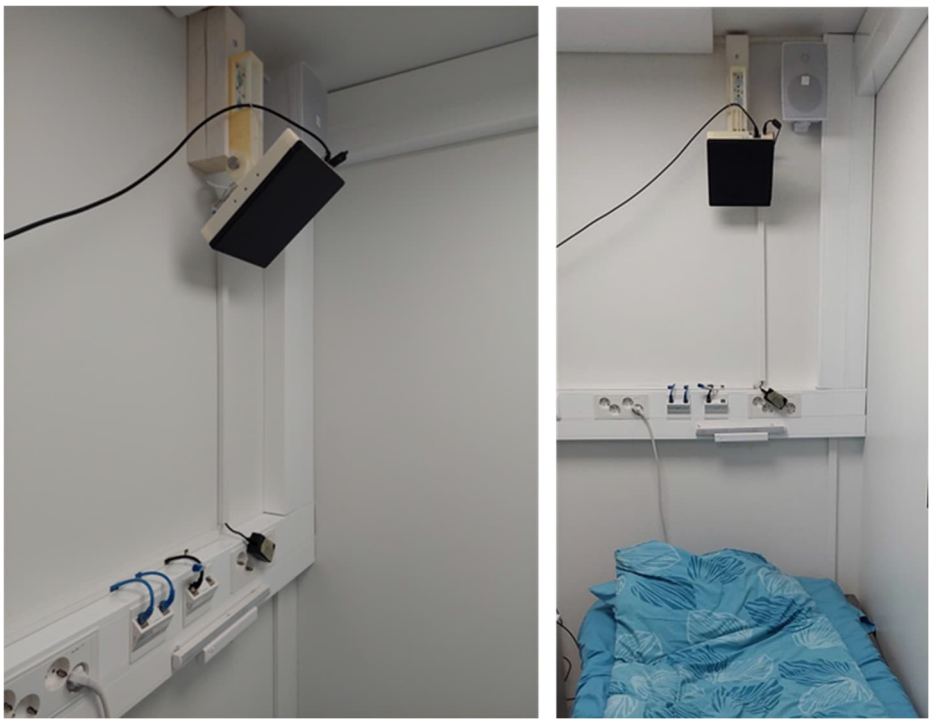


Figure 25 VTT radar mounting on the wall above the bed at P1-KSL. The radar is tilted downwards to focus on the bed.

3.5.2 Pilot evaluation execution and protocol details

This section describes the Finapres blood pressure monitoring and the Bittium HSAT set. The VTT Radar, eLive eBedSensor and Edge system has been already described in the previous section of P1-KMPHG.

3.5.2.1 Finapres CNIBP monitor

Finapres develops a Continuous Non-Invasive Blood Pressure monitoring system. In the DistriMuSe project, a wireless variant is developed, however this is not finalized before the first testing cycle. A Finapres NOVA is utilized during the first testing cycle of the project. A low impact on the use case is expected, since measurement principle is the same. The main unit in this case is not portable, not allowing patients to walk around with the device. As there are also non portable other systems attached, this is not deemed to be a problem. The system consists of a main unit, providing the relevant GUI and Analog IO options, see the next Figure. The main unit is connected to a NANO CORE, similar to what will be used in the wireless system. This will allow the pilot to build experience with the system without having the finalized system available.

Figure 26 Finapres NOVA

The NOVA will be used to monitor blood pressure of the subject continuously during the sleep study, using the Volume Clamp method. Signals can be synchronized in post processing with other measurement signals using the analogue input signal as a time reference.

3.5.2.2 Bittium HSAT set

Bittium Home Sleep Apnoea Testing solution (HSAT) is an impressive combination of a measuring device, analysis software and service platform.

Figure 27 Bittium HSAT

Bittium aims to study less obstructive methods to monitor sleep time and sleep stages using multi-sensor fusion from non-EEG signals that are commonly used in the home sleep apnoea tests. Used signals include PPG, respiratory effort, respiratory airflow and acceleration.

The data acquired in the P1-KSL pilot will be used in development of sleep time and sleep staging analysis for Bittium Respiro. The gold standard PSG data will provide reference for evaluating the developed algorithms.

3.5.3 Pilot technical KPI measurements

Firstly, the Finapres NOVA system was tested at VTT Oulu facilities. The feasibility of the device for overnight CNIBP measurement was evaluated with the following protocol:
1) Measurement of the CNIBP with single cuff
2) Measurement of CNIBP with alternating double-cuff with 10 minutes switch interval
3) Alternating double-cuff with 5 minutes interval

This was executed for three different persons to assess the possible pain/inconvenience produced by the finger cuffs. With single cuff, 10-15 minutes is maximum length of measurement without major inconvenience. With alternating cuffs, both 5 and 10-minute switch intervals are usable while awake. For sleep recording usage, we concluded that 10-minutes would be the best trade-off between pressure-related inconvenience and sleep disturbance.

Second, the Bittium system was tested at the SmartSleep Laboratory in May 2025. In addition to other signals, it measures ECG and pulse rate via oximetry. The comparison of heart rate evaluations between the Bittium and other sensors is shown in Figure 28.

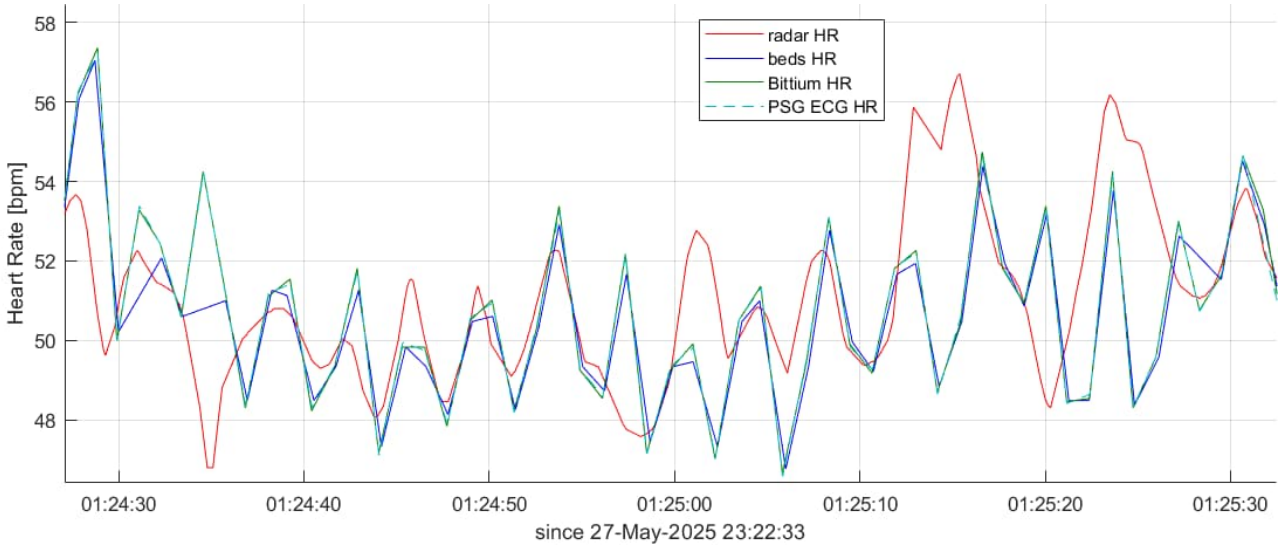


Figure 29 Heart rate beat-to-beat values from the P1-KSL pilot. VTT radar HR is with red color, eBedSensor with blue color, Bittium with green and the reference PSG HR is with light blue dashed

The HR calculated from the Bittium ECG and eBedSensor accurately resemble the gold standard. VTT radar HR has difficulties in detecting HR accurately in this example. In the previous studies, the interpersonal variation between the study subjects has shown high variance on the accuracy of the sensors. Statistics for the measurement accuracy shall be reported for the final P1-KSL pilot data.

One-minute averaged HR showed better accuracy for all devices. In the example figure 30, the VTT radar with red color had problems in detecting the sudden increase in the one-minute averaged HR. This is a noteworthy shortcoming, as in sleep registrations the rapid heart rate swings are common, and the timescale is in seconds rather than minutes.

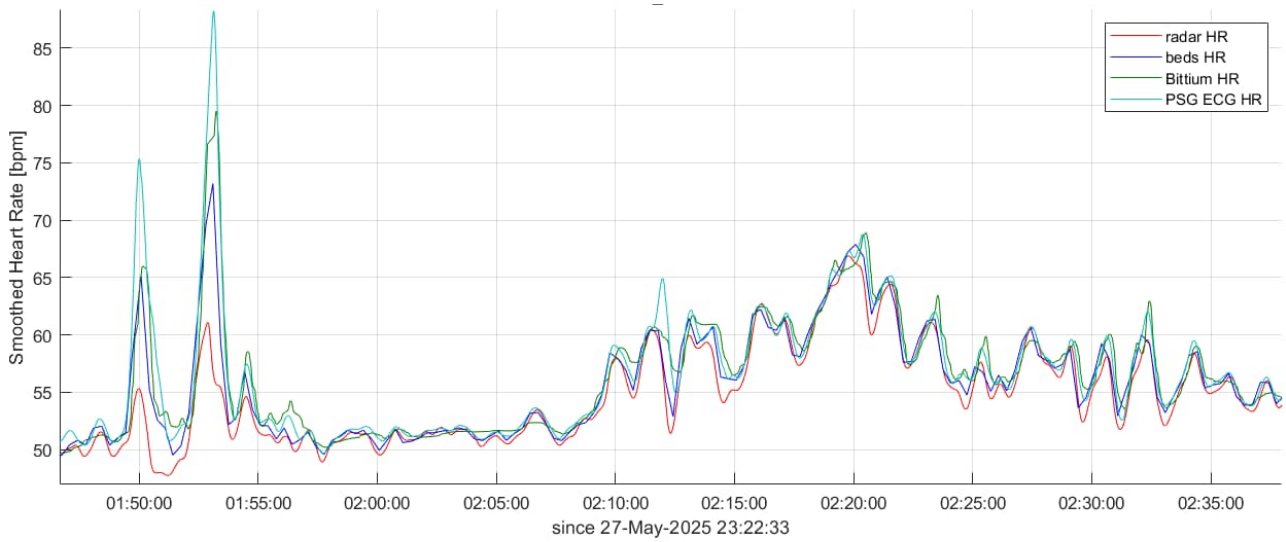


Figure 30 One-minute averaged heart rate measurements from the P1-KSL pilot. VTT radar HR is with red color, eBedSensor with blue color, Bittium with green and the reference PSG HR is with light blue

Time synchronization between the sensor recordings is essential. The detected time offset between the eLive Edge and the PSG system for two separate nights is presented in figure 31.

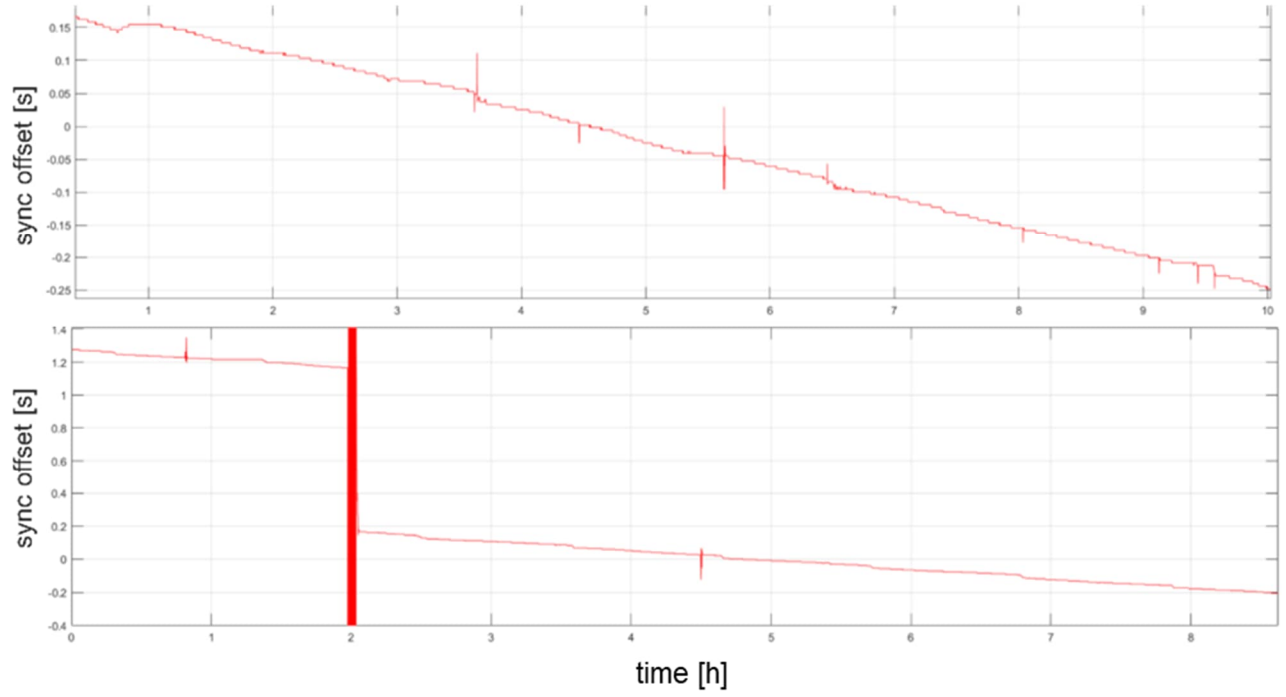


Figure 31 Time offset between the eLive Edge system and the PSG during two separate night recordings. The vertical axis is the detected time difference in seconds. The horizontal axis is hours from the start of the PSG recording. The first recording is about ten hours and the second recording is about eight hours long.

The detected time offset drifts approximately 0.5 seconds in ten hours. There is also a sharp step of around one second in the other recording at 2 hours from the beginning of the recording (Figure 31, bottom graph). Both behaviours were as expected; the drift between different system clocks is

typically about one second for 24 hours. Moreover, the PSG systems internet based NTP synchronization for the so-called wall-clock time causes occasional step-like corrections. The eLive Edge recording uses the monotonic CPU time which is not connected to internet NTP synchronization; only the measurement start time is taken from the wall-clock time.

Occasional peaks in the detected offset difference are probably outliers caused by the cross-correlation method, as it is expected that the offset difference between different system clocks shall either drift monotonically or have step like changes. We removed these outliers with median filtering before further processing steps.

The time offset between the two recordings of the external pulse signals was derived with a sliding window cross-correlation method. The external sync pulse signals before and after fixing of the time offset is presented in figure 32.

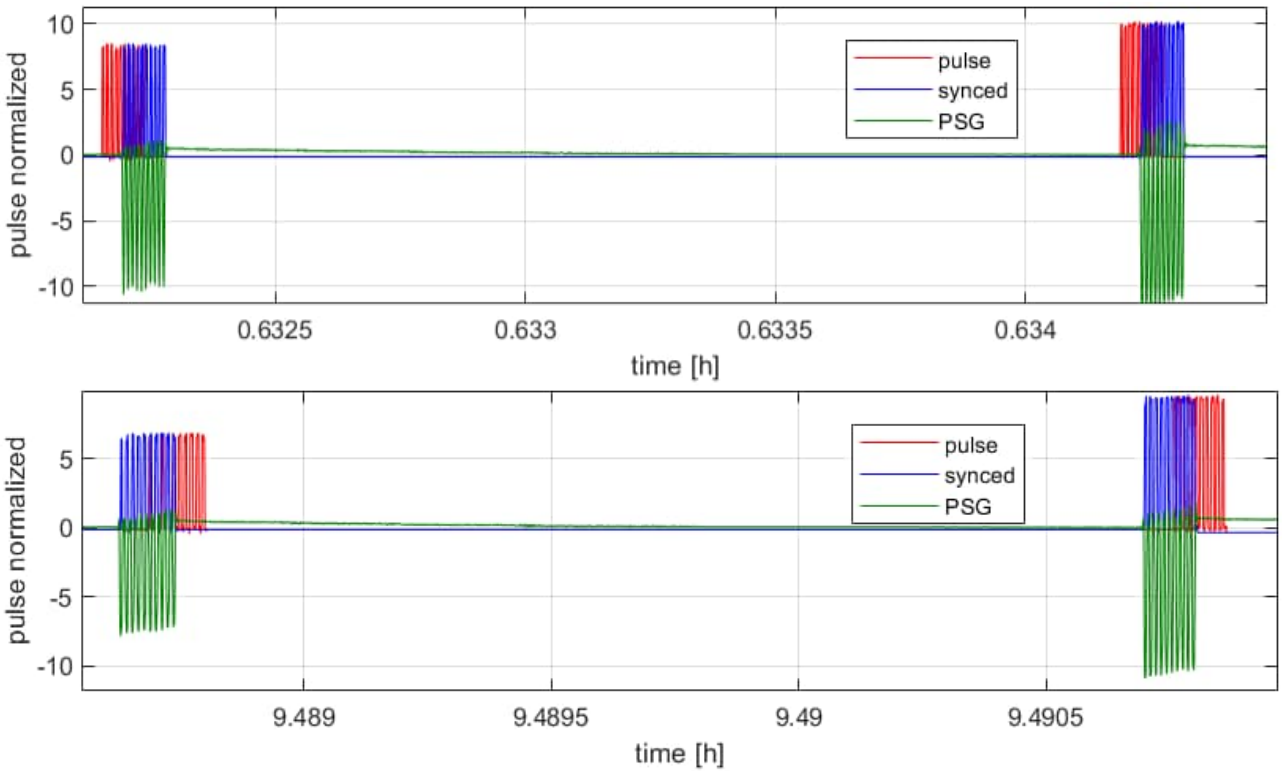


Figure 32 Pulse sync signals at early night and at late night of the first recording from the figure 31. The red curve is the original pulse signal from the eLive Edge, the green is the (negative) pulse signal measured with PSG, and the blue curve is the offset corrected Edge pulse signal. The horizontal axis is time in hours as corresponding with the previous Figure. The horizontal grid width is 1.8 seconds long and the length of the zoomed period is about ten seconds

The top graph in figure x shows a time offset of about 0.15 seconds. The original eLive Edge pulse signal timestamps (red) were 0.15 seconds before the corresponding PSG recording (green). The bottom graph in figure x shows a –0.25 seconds time offset. The offset corrected eLive Edge pulse signal is shifted with the cross-correlation method to match accurately with the PSG timestamps. The same offset shift is applied to all eLive Edge recordings.

eLive eBedSensor HR beat-to-beat value against the reference PSG is presented in figure 33.

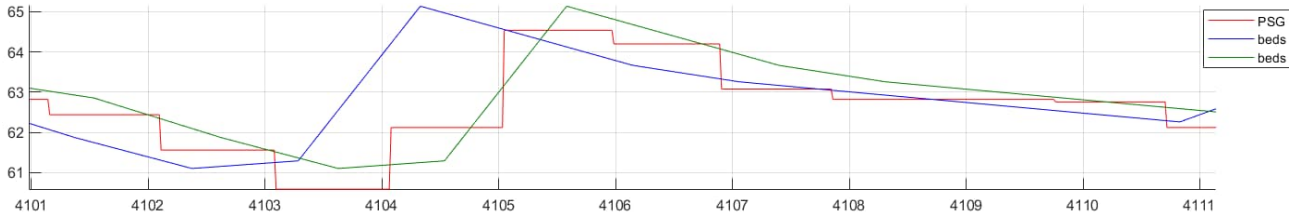


Figure 33 Red curve shows the PSG ECG based beat-to-beat HR values and the eBedSensor HR values before (blue) and after time offset correction (green). Vertical axis is heart rate [bpm] and the horizontal axis is time in seconds.

The eBedSensor HR match well with the reference PSG after fixing the time offset.

3.5.4 User aspects: stakeholder engagement in pilot development

Stakeholders of this pilot (eLive, VTT, Bittium, Finapres) have participated in the planning and setting up the study environment, providing equipment and devices, detailing the data collection protocol among other technical aspects. In this pilot target users are not engaged as they are in P1-KMPHG.

3.5.5 User-based KPI assessment

The user-based KPI are the same with the both sleep laboratory pilots and they have been described in the previous section of the P1-KMPHG in the corresponding KPI table.

3.5.6 User aspects: Gender/age issues and ethical concerns in P1-KSL

This pilot did ethics focus group on 16.5.2025. Moreover, favourable ethical statement for the pilot study protocol has been received from the Regional Medical Research Ethics Committee of Eastern Finland Collaborative Area on 19.8.2025.

3.5.7 P1-KSL transition into cycle 2: takeaways and feedback

During cycle 1, we managed to do test measurements for all the devices included in the pilot. However, most of the work has been preparatory work to conduct the actual pilot measurements in cycle 2. This is due to Business Finland´s delayed grant decision postponing the start from 1.5.2024 to 1.3.2025, finishing the paperwork related for favorable ethical statement and the paperwork needed for device investigation permit from Finnish Medicines Agency. No changes have been planned for cycle 2.

3.6 P1-BRNO evaluation cycle 1

3.6.1 Final pilot set-up

Pilot P1-BRNO, coordinated by Brno University of Technology (BUT), aims to validate a comprehensive, multi-sensor wearable platform in realistic free-living and sports environments. The solution includes ECG monitoring, continuous glucose monitoring (CGM), sweat sensing, and physical activity tracking, combined with a mobile application for data feedback. The pilot investigates how such a system can be used to provide real-time, user-friendly feedback to athletes

and their coaches, with the ultimate goal of supporting performance enhancement and reducing health risks associated with overtraining, inadequate nutrition, or improper recovery.

The pilot is designed in two cycles. In Cycle 1, the focus is on evaluating the feasibility of the envisioned demonstrator and collecting preliminary user feedback and physiological data in the field. Using the data, algorithms for user feedback will be developed and integrated in mobile application in Cycle 2. This phase will provide critical input and necessary algorithms for Cycle 2, which will incorporate technological and design improvements based on user input. The second cycle will validate the improved solution with a higher level of technological readiness and usability, supporting the pathway toward future commercialization.

The pilot targets recreational and semi-professional athletes as primary users. The pilot is conducted by the Centre of Sports Activities in collaboration with the Department of Biomedical Engineering at Brno University of Technology. Testing was conducted in free-living and sports training conditions, with attention given both to human-centred aspects, such as system usability, comfort, feedback quality, and user acceptance and to technical KPIs relevant for the system's overall performance and reliability.

3.6.2 Final pilot set-up

The pilot set-up was not changed since D5.1 and D5.2. Pilot evaluation execution and protocol details. The whole pilot architecture (cycle 1 and cycle 2) is briefly described and repeated from D5.1.

The pilot architecture consists of 4 layers (not all of them are addressed in cycle 1). The pilot architecture overview can be seen in Figure 34. In the first layer there are various commercial and custom-made wearables. The data from wearables are transmitted via Bluetooth (using BLE GATT or custom protocols) to gateway (a smartphone was used for testing), which is the second layer. The data are preprocessed on this edge device into CSV file format and transmitted to the third layer which is cloud. On cloud, demanding computation tasks should be performed (for testing purposes, computational-powered work-stations were used). The outputs are visualized in the fourth layer to give the user feedback via custom-made smartphone application.

The first layer is the sensor layer of wearable sensing devices attached to the body of the participant. The sensor layer in cycle 1 includes glycaemia sensor (a wearable CGM sensor) and chest sensor (a wearable device with integrated ECG and ACC sensors) and commercial smartwatch for activity tracking.

The second layer is data gateway and sensor interconnection. Sensors are connected to the edge device using a Wearable Bluetooth Network (WBAN) which combines BLE GATT protocol and custom protocols for selected devices (Faros, Smartwatch). In the first cycle, the data was stored on the edge device for development of processing algorithms such as detection of QRS complexes in ECG or calculation of activity type. An energy efficient signal parametrization algorithm will be also implemented. This algorithm will calculate statistical features that are going to be necessary for activity classification and sleep detection. In the next phase, the data should be pre-processed on the edge device directly using the developed algorithms. From the edge device, the data is later relayed to the cloud. The edge device also serves as a preliminary data storage.

The third layer of pilot architecture is cloud computing. In cycle 1, we have measured data essential for development of algorithms which will be implemented in cloud in cycle 2 and will give the feedback

to the user (cardiac pathology types and severity, energy expenditure, user exertion, glycaemia variability, and health score). The activity classification and sleep detection from extracted features will be done in this layer.

The last layer focuses on the visualisation of the data to the user. In cycle 1, the user sees only a portion of the data provided by the app of the commercial developers.

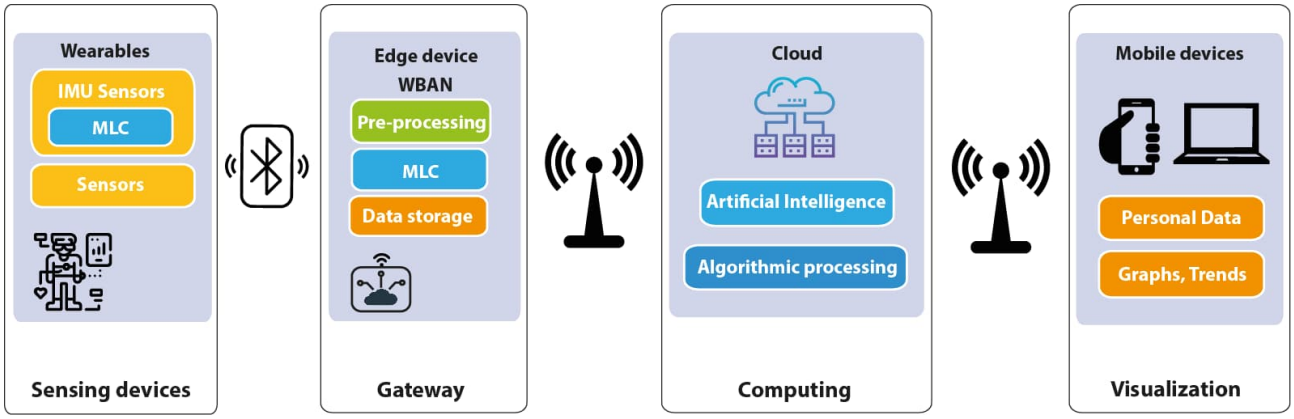


Figure 34 P1-BRNO architecture.

3.6.3 Pilot technical KPI measurements

With this demonstrator, we aim to validate a set of technical KPIs that reflect the progress and innovation achieved within the pilot activities. The pilot-level KPIs have been defined per component and functionality, ranging from integration of novel sensors, data transmission, signal processing to health-related analytics. Each KPI has a clear target value or qualitative objective to be reached. The following Table *11* outlines the baseline state and expected target values for each KPI.

Table 11: Technical KPIs for P1-BRNO

Component, KPI	Start state	Target
KPI 2.1: Early in-sensor fusion Decreasing of data rate between IMU/wearables and edge gateway with simultaneous maintaining battery capacity.	Current energy consumption with raw data transfer from IMU/wearable is more than 5 times higher with lack of any processing.	Energy consumption of IMU chip/ wearable max. +30% while decreasing data transfer rate by a factor 5 and data fusion using edge computing.
KPI 3.2.1: Distributed computing Decrease of traffic transmission in the system.	Baseline is a centralized approach (absolute volume depends on sensors, e.g. 5–30 kbit/s for ECG sensors).	At least 50% less traffic transmission compared to a centralized approach.
KPI 4.5: ECG pathologies detection ECG pathologies detection integrated within standard parameters evaluated by wearables.	Independent system	Integrated system
KPI 4.6: Health assessment Health assessment will be validated at least on a few subjects and compared with their subjective and/or objective diagnoses and feelings.	Health score in % based on physical condition	Complex health assessment based on many parameters provided by wearables including new integrated sweat sensors
KPI 4.9: Smartwatch with integrated sweat sensors		
Wireless and battery operated	BLE wireless technology, 10-hour operating time	24-hour operating time
Number of ML core chips Advanced perspiration rate sensor	0	1-3

Environmental compensation for perspiration rate (PR) measurement	Basic PR determination	Compensation to environment
Sensor integration	Independent device	Integrated device
Edge Device		
Edge Device battery lifetime	N/O	4 hours
Deep discharge battery protection	N/O	Present
Number of simultaneously connected sensors	1	5
Reliable Bluetooth LE communication with sensors	N/O	Present
GPS position accuracy	N/O	Max 10 m

Energy consumption reduction – this KPI from D5.1 has been thoroughly investigated as low energy consumption is crucial for a battery-powered device. For low energy consumption statistical feature calculation was chosen as the feature processing algorithm. These measurements were conducted on MAX78002EVKIT development board as this board used the MAX78002 microcontroller which is going to be used on the final edge device.

This board also contains a PMON display which can be used for simple and effective measuring of energy consumption. The system behind PMON display can be configured to measure only specific parts of code within the chip, this was used to measure only the consumption of the feature extraction algorithm. The method of measurement was as follows – first measurement was taken of the chip without any algorithm running on it then the algorithm was implemented, and another measurement was done. These 2 values were then subtracted which led to the result representing only the energy consumption of the feature extraction algorithm.

To account for potential fluctuations in the chip's energy consumption – arising from both internal and external influences – each experimental configuration was measured 20 times. This repetition aimed to enhance the statistical robustness of the results and minimize the impact of random deviations. The arithmetic mean of these 20 measurements was then calculated and used as the representative value for the chip’s behaviour under each specific condition.

For reference a Convolutional neural network (CNN) was trained that did both feature extraction and activity classification on the chip. This represented the classic approach of activity recognition using wearable devices that is common in many technological solutions. Results can be seen in Table *12*.

Table 12: Energy Consumption and Computation Time

Type of Algorithm	Energy Consumption [µWs]	Computation Time [ms]
Statistical features	3417.25	972.9
Statistical features (no spectral features	483.3	304.0
CNN	18686.6	2017.1

The results demonstrate that usage of edge device for statistical feature extraction with server for final classification reduces energy consumption by approximately 5.5-times, which prolongs the lifetime of the battery-powered edge device. In the scenario of avoiding the use of spectral features, the energy consumption becomes even smaller. This can be attributed to the high energy usage of

the FFT algorithm. Despite this avoiding the use of frequency-based has a significantly negative impact on the accuracy of classification models. In Table *13* we compare the best model (random forest – RF from D5.2) with and without usage of spectral features.

Table 13: Comparison of the best models with respect to the use of spectral functions

Model	Accuracy	Precision	F1 Score	Recall
RF without spect. Feat.	91.85%	91.31%	91.15%	91.08%
RF with spect. feat.	96.85%	97.21%	97.10%	97.07%

The observed data clearly indicate a decline in classification performance metrics under both evaluated scenarios. Considering that the incorporation of spectral features yields a 5.5-fold reduction in energy consumption, and that their omission leads to a substantial degradation in classification accuracy, it can be concluded that the inclusion of spectral features is essential for achieving optimal model performance.

Traffic transmission reduction – as an important factor, the data volume of statistical features was measured. Measurements were conducted again on MAX78002EVKIT. The board was connected to a PC via UART over USB. Consequently, communication was carried out using the UART protocol. A Python script utilizing the *pyserial* library was developed to monitor the designated serial port over a fixed time interval. During each 10-second period, the script measured the cumulative amount of transmitted data. This method was used to compare 2 stats – transmitted data volume when calculating statistical features and transmitted data volume when sending raw data without any processing. The results can be seen in Table *14*.

Table 14: Transmitted Data Volume

Type of Data	Transmitted Data Volume [B]
Statistical Features	170.1
Raw Data	6874.7

These results demonstrate that usage of statistical features calculation as a method of feature extraction managed to decrease the necessary volume of transmitted data approximately 30-times, which follows the related KPI from D5.1.

ECG pathologies detection – ECG pathologies detection is highly dependent on the dataset with annotated pathologies in ECG. Therefore, the main focus was on obtaining such a dataset and reading the data in raw format. The algorithm for deciphering data from Custo Diagnostic was successfully implemented and, the dataset containing approximately 7500 ECG recordings from the sports doctors‘ laboratory was obtained. In addition, a Python script for communication with Polar H10 (single-lead wearable ECG device) in real time was devised using *bleak* and *bleakhearth* libraries.

Health assessment – Our health assessment algorithm is based on a combination of parameters that can be obtained from a questionnaire or measured using wearable devices, such as age, sex, height, weight, waist circumference, hip circumference, smoking, alcohol consumption, resting heart rate, heart rate variability, blood pressure, blood glucose, and cholesterol. We are now extending our algorithm with the so-called cardiovascular age. By merging several publicly available databases, we have created a dataset containing records from more than 100,000 people with reference values. The dataset includes signals from both healthy and nonhealthy individuals. Using this dataset, we train the algorithm to determine cardiovascular age. At present, we are working to expand our dataset with additional data to test the algorithm on independent samples.

Smartwatch - wireless and battery operated – the Smartwatch device integrates sensors of IMU (accelerometer and gyroscope, perspiration rate, and temperature). It enables Bluetooth communication with a data concentrator. The transmitted data are in raw format as for the first phase, then will be processed by ML-core and transmitted in a reduced form on demand from the planned edge device. The data transmission was validated using a smartphone app which can connect to the sensor and monitor the received data. The Smartwatch was tested for 24h operating time as shows Table *15*. During the operating time, the device measured data using sensors and Bluetooth connection was established at check time.

Table 15: Operating time

Testing action	Datetime
Smartwatch start	29.7.2025 9:02
Check Smartwatch is running	30.7.2025 12:08

Salinity sensor integration – in the first phase, the small size standalone mobile conductivity sensor with small cell was integrated to android application where the measured data is sent in using BLE (Bluetooth Low Energy). The integration was tested with a smartphone app which connects to the sensor and listen for incoming data that are shown in a graph on Figure 35. The 3-point calibration of the sensor was used. The communication is achieved using standard GATT technology and the specific conductivity value in mS/cm is updated once per 2 seconds. The salinity is calculated from the conductivity within the normal content of substances in sweat. Normal specific conductivity of sweat is in the range of 10.9 -11.6 mS/cm. The sensor has a range of 3 – 120 mS/cm.

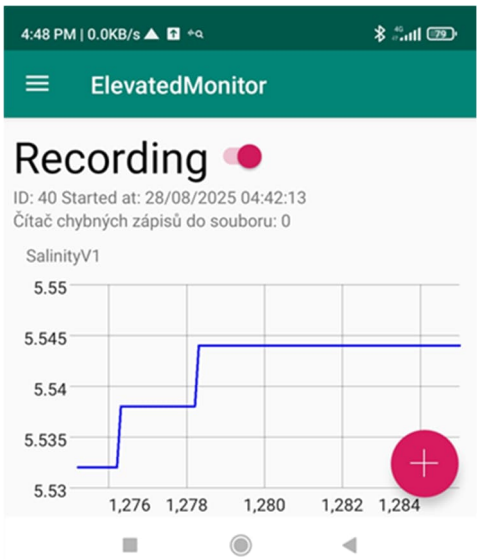


Figure 35 Real-time specific conductivity data (in mS/cm) from Salinity sensor via BLE

Specific conductivity was measured using a sensor developed within the project. Sodium chloride solutions of known salinity were prepared and tested. Preliminary experiments demonstrated good reproducibility of the conductivity measurements.

Based on the measured data (Figure 36) the following power-law relationship between salinity and specific conductivity is concluded:

$$Specific\ conductivity = 2.8001 \cdot (Salinity)^{0.862},$$

with R^2 value of 99.96%.

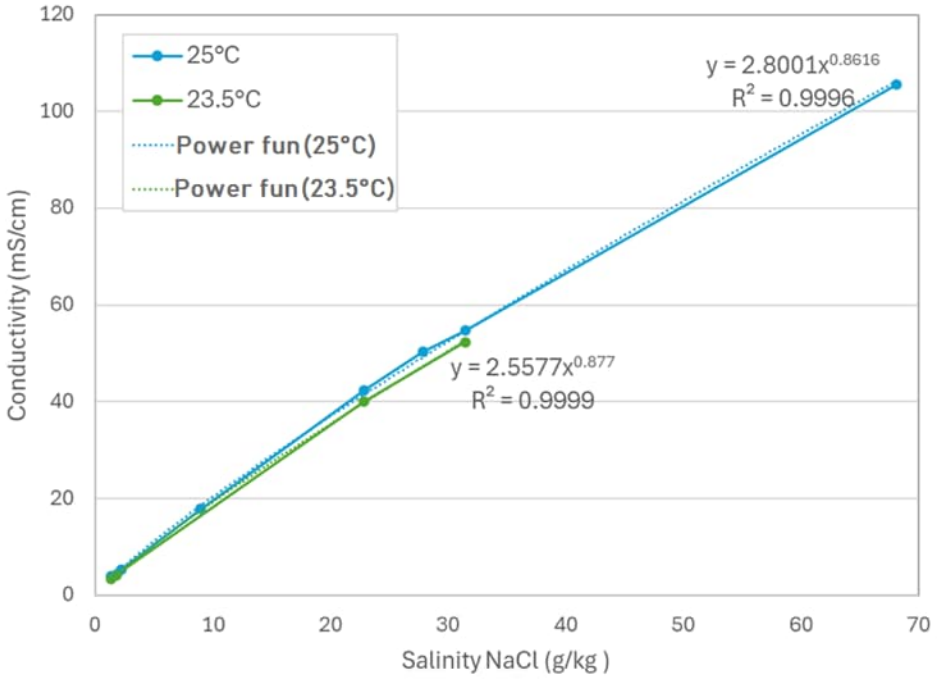


Figure 36 Specific conductivity (mS·cm−1) vs salinity (g·kg−1) of sodium chloride at various temperatures.

Edge Device battery lifetime - The Edge Device is battery operated so the lifetime after the battery is fully charged is quite important. The device consists of three basic components with the highest energy consumption. They are the main processor unit, LTE modem and ML processor. For the first period consumption without the ML processor has been evaluated. Measurement with all components is planned for the second project period.

For the test the LiPol battery with capacity of 6,000mAh has been used. The battery was fully charged at the beginning of the test. All components have been active. The test was stopped when the output battery voltage dropped to the value of 3.0V. Table 16 details the measurement result.

Table 16: Time measurement

Test start time [h:m]	Test end time [h:m]	Duration [h:m]
8:15	12:32	4:17

The Edge Device lifetime overcome the desired time. On the other hand, the consumption with ML processor will be higher, so in the project second period some consumption optimisation will be necessary.

Deep discharge battery protection - For Edge Device durability it is necessary to take good care of the device battery. Besides the proper battery charging is a big importance to avoid the situation with deep battery discharge. It can happen, when the load is present even the battery is discharged. In this case the battery could be damaged. To avoid this situation the Edge Device is equipped with

circuit, which cut off the load in case the battery is discharged. For this situation is typical the output voltage drops to 3.0 V and less. The test verified, that the load was cut off when the output voltage from battery dropped to critical value. After the charging the battery the Edge device has been operational again.

Number of simultaneously connected sensors – The Edge Device is designed to collect data from several body sensors. The number of sensors will increase in the future with the number of measured values. Therefore, the Edge Device has to communicate at one time with more sensors. So, for the test five BLE devices have been chosen. They are two heart rate wrists monitor MIO and three heart rate chest monitors Polar H10. The data have been sent in 1s intervals and have contained value for heart rate and inter beat intervals (Polar). All data was logged and evaluated at the end of the test. All data from all five sensors have been successfully captured. So, the criterium was met. Moreover, the test proved another KPI - **Reliable BLE communication with sensors**.

GPS position accuracy – the position of Edge Device is measured by uBlox module MAX M10s with external flat antenna. The position is based on GPS satellites so only outdoor position can be obtained. The goal for measurement is to reach better accuracy than 10 meters. The module sends the position in degrees. The measurement has been done in Prague, where N10E-5˚ correspond to 1.11 m and E10E-5˚ corresponds to 0.715 m. The position of the defined place was N50,067314˚ E14,410125˚. Table 17 contains maximum and minimum measured values.

Table 17: Position deviation calculation

Item	Maximum	Minimum	Maximum distance from the defined position [m]
Nord	50.067335˚	50.067291˚	2.55
East	14.410147˚	14.410101˚	2.66

The maximum deviation is about 2.66 m, so the measurement accuracy is better than the desired 10 m.

3.6.4 User aspects: stakeholder engagement in pilot development

During the first cycle of the P1-BRNO pilot, we identified and actively engaged various stakeholder groups and end-users to ensure the system development aligns with their needs and expectations.

- Recreational and semi-professional athletes, who are also students or employees of a university, aged 18–50 years, male and female, participating in a ten-day free-living monitoring period as well as laboratory tests.
- Sports coaches and healthcare professionals who contributed to defining requirements and setting up testing protocols.
- Healthcare professionals (Holter technicians), consulted regarding the correct use and integration of measuring devices.

Stakeholder involvement was conducted primarily through individual qualitative interviews with selected participants after the completion of free-living monitoring and laboratory tests. These interviews aimed to gain deep insights into user needs and preferences. Additional consultations

were held with coaches to gather their perspectives after their athletes' completed measurements, and healthcare professionals provided advice on improving comfort and proper use of Holter devices.

The engagement goals were to verify the usability and acceptability of wearable sensors and related applications, identify key user needs and aspects requiring improvement for the KPIs, and prepare materials for creating high-quality quantitative questionnaires for broader evaluation in Cycle 2. Moreover, the engagement helped to collect valuable data supporting further development and innovation of the pilot system, as well as to gain new insights into physiological functioning.

This first phase of stakeholder engagement took place after the ten-day monitoring and laboratory testing period, with further involvement planned for Cycle 2 based on the insights obtained.

3.6.5 User-based KPI assessment

Based on qualitative interviews with selected participants, we obtained initial findings regarding the user-based KPIs previously defined. The methodology involved in-depth interviews with selected pilot participants after the ten-day monitoring period, followed by analysis focused on wearing comfort, device usability, clarity of data presentation, and perceived value of the feedback provided.

These qualitative insights informed the development and refinement of quantitative questionnaires that will be distributed to all participants in Cycle 2 to allow broader and more robust KPI evaluation.

The KPIs assessed include usability rating (SUS), wearability score, perceived accuracy, perceived usefulness of feedback, willingness to adopt the solution, and the number of users involved in the pilot testing.

Preliminary findings highlighted common issues such as:

- The most frequent issue was skin irritation caused by Holter electrodes during prolonged use.
- Users missed a unified platform for integrating and visualizing all measured data and expressed dissatisfaction with having to use multiple separate applications, which complicated data interpretation in a broader context.
- Users lacked comprehensive feedback including practical recommendations, such as alerts about unsuitable diet leading to significant glycemic fluctuations or suggestions to adjust meals prior to physical activity for improved performance.

Planned activities for Cycle 2 include quantitative KPI evaluation on a user sample based on the refined questionnaires, testing alternative electrodes of ECG recorders and different types of ECG recorders aimed at enhancing wearing comfort of ECG recorders, and developing a centralized application that integrates data from all sensors to improve the overall user experience.

3.6.6 User aspects: Gender/age issues and ethical concerns in P1-BRNO

The project met its ethical requirements. A total of 50 participants were enrolled, and informed consent was obtained from each participant prior to any measurements. The gender balance was skewed toward men, therefore, more women will be deliberately included in the next phase. The age distribution had a majority of participants under 30; accordingly, recruitment will be expanded to include more participants aged 30–50.

A focus group focused on ethics was conducted using the DistriMuSe Ethics Exercise Tool. Eight people involved in the research participated in the group, of whom 6 also participated in the testing

(target users). It was not possible to include test participants who were not part of the research team because the testing ended much earlier than this tool was available. The results will be reported in D7.7. The results will be considered for next pilot evaluation in 2nd Cycle.

3.6.7 P1-BRNO transition into cycle 2: takeaways and feedback

The pilot aims to validate a wearable, multi-sensor platform for athletes and their coaches. The goal was to enhance performance and mitigate health risks by providing real-time feedback. **Pilot Goal** was to validate a wearable system with sensors for ECG, continuous glucose, sweat, and physical activity. The system provides real-time feedback through a mobile app to help athletes and coaches with performance, overtraining, and recovery. Cycle 1 focused on testing the demonstrator's feasibility and collecting preliminary user feedback and physiological data. This data will be used to develop algorithms and make technological improvements for Cycle 2, which will focus on validating the refined solution for potential commercialization. There were **technical achievements** such as a four-layer architecture using wearable sensors, a smartphone as a data gateway, cloud computing for complex analysis, and a mobile app for user feedback; **energy and data reduction** by using statistical features on the edge device decreasing energy consumption by approximately 5.5 times and data transmission volume by 30 times compared to sending raw data; and **health analytics** based on Datasets that were acquired to develop algorithms for ECG pathology detection and a "cardiovascular age" health assessment. The ethic issues were considered from **user feedback** that was evaluated through qualitative interviews, regarding to **comfort** due to participants experienced skin irritation from Holter electrodes, regarding to **usability** because users were dissatisfied with having to use multiple separate apps and wanted a single, unified platform for all their data and finally users desired more comprehensive and practical feedback, such as dietary suggestions or alerts.

Plans for Cycle 2 is Based on the Cycle 1 findings, where the team plans to conduct a quantitative evaluation with a larger user sample, test different types of ECG recorders and electrodes to improve comfort, develop a centralized mobile application to integrate all sensor data, improving the overall user experience, and expand recruitment to include more female and older participants (30-50 years old).

3.7 P1-TOR evaluation cycle 1

3.7.1 Final pilot set-up

The pilot P1-TOR is conducted by University of Turin (UNITO) in collaboration with University of Parma (UNIPR) and aims at developing a multi-sensor platform designated to monitor motion in two main contexts: (a) sport performance assessment in non-professional athletes and (b) health monitoring of frail individuals, including elderly healthy people and subjects affected by Parkinson's Disease (PD). The pilot activities are conducted at the "Istituto Auxologico Italiano" in Piancavallo, Italy, and at the scientific campus of UNIPR, Italy, where PD patients and non-professional athletes, are being respectively recruited. Motion-related data of interest are collected according to a common protocol, specifically designed by physicians from UNITO, consisting in a set of medical tasks, typically employed to evaluate the severity of PD patients by means of the Unified Parkinson's Disease Rating Scale (UPDRS).

The system set-up developed to acquire motion data is composed by heterogeneous commercial sensing devices, including (a) Garmin smartwatches to collect physiological parameters in addition to motion information on the users' daily life, and (b) IMU-based devices to collect motion features related to the users' gait. To this end, the use of different IMUs is being tested to identify the least invasive set-up which ensures a good data accuracy. In an initial evaluation phase, the custom-made move2i® device was placed at the lumbar region of the users and was used in combination with the full XSens MVN Awinda motion capture system, adopted as benchmark. At present, the move2i® device is being employed together with a single node from the full XSens system (XSens MT node or XSens Movella Dot), for further validation of the custom-made device. Both sensors are secured at the L5/sacral level by means of elastic fabric belts. The final set-up therefore consists of a single-node IMU positioned on the lower back of the users. Figure 34 illustrates the full set-up worn during the evaluation phase; the two nodes employed in the final set-up configuration are highlighted in the dedicated inset.

On the other hand, Garmin smartwatches are being employed to collect physiological and motion data from non-professional athletes during their daily life activities. Monitoring healthy subjects with these devices has led to the development of an innovative metric, the Physical Activity Index (PAI), designed to quantify the physical activity of the monitored individuals. A small group of PD patients will be equipped with Garmin devices to evaluate the PAI, also under clinical conditions.

All the components employed in P1-TOR are shown in Figure 37.

Figure 37 Full system configuration and inset of the final set-up

Figure 38 Components used in P1-TOR

3.7.2 Pilot evaluation execution and protocol details

Within the pilot P1-TOR, aspects concerning the invasiveness of the set-up and the implementation of the data acquisition protocol are currently under evaluation. The protocol designed by UNITO is structured in 4 acquisition phases:

- Preliminary evaluation phase;
- “in-lab” phase;
- “out-of-lab" phase;
- “long monitoring” phase.

During the first preliminary evaluation phase, participants are recruited by investigators who are in charge of illustrating the benefits and risks of participating in the pilot. The investigators will also collect the written informed consent duly completed and will be responsible of the data collection campaign, which will take place in laboratory and domestic settings.

During the “in-lab” phase, a set of structured exercises, illustrated in Figure 39, is performed in order to evaluate the severity of the PD. Each task is briefly detailed hereafter.

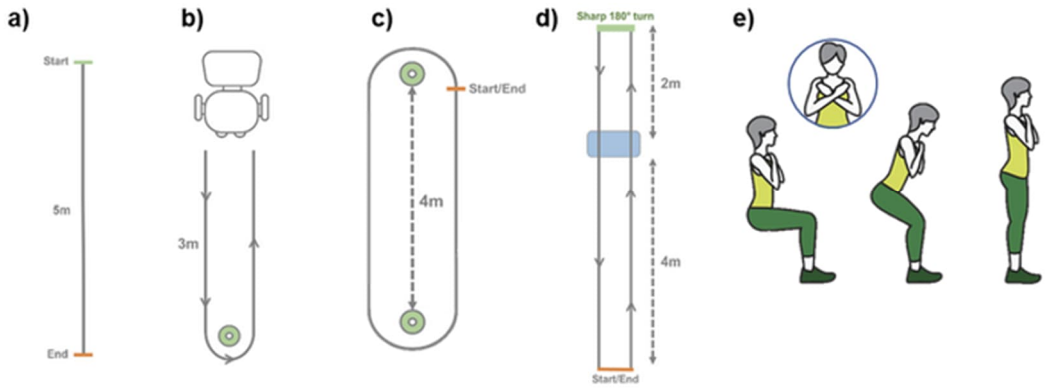


Figure 39 Data collection protocol used during the “in-lab” phase

Straight line Walking (SW). The participant walks along a predefined path of six meters from a starting point to a final point. This exercise is performed twice at three speeds: comfortable, slow and fast.

Timed Up and Go (TUG). The participant stands up from a chair and walks three meters to a cone, then turns around it 180° and walks back to the starting point where they will sit down. This is done at a comfortable speed.

Two Minute Walking Test (2MWT). The participant walks at a comfortable speed around a circuit around two cones spaced six meters apart for two minutes at a time. After the time is up, the participant returns to the starting point.

Hallway Test (HT). The participant stands at a starting point and walks at a comfortable speed for four meters to a step that they will climb up and down, then walks another two meters, turns around a cone and runs the circuit in reverse to the starting point.

30-Second Chair Stand (30CST). The participant stands up from a chair with their arms crossed at the wrists. Both feet are kept in contact with the ground. The exercise is repeated for 30 seconds or less if the participant is unable to.

In the current phase of pilot evaluation, the SW test is being assessed to test the system set-up in terms of accuracy of the collected data. Th SW is also common to TUG, 2MWT and HT tests, where participants are required to walk along a straight line along specific paths. These more complex tests will be assessed during the Cycle #2, when also the “out-of-lab” and “long monitoring” phases will be carried out by continuously monitoring the participants during their daily life activities.

3.7.3 Pilot technical KPI measurements

Within the pilot P1-TOR, two main technical KPIs and their respective evaluation criteria have been identified and are reported in Table 18.

Table 18: technical KPIs for pilot P1-TOR

Component, KPI	Start state	Target
KPI 3.2.1: Distributed computing Traffic transmission amount decrease in the system	Baseline is a centralized approach (absolute volume depends on the considered sensors)	At least 20% less traffic transmission compared to a centralized approach
KPI 4.6: Health assessment Health assessment will be validated at least on a few subjects and compared with their subjective and/or objective diagnoses and feelings	Health score in % based on the physical condition	Complex health assessment based on multiple parameters provided by wearables

1. **KPI 3.2.1 (Distributed computing)** aims at improving the data transmission efficiency by reducing the communication traffic through distributed processing at the edge node level. In particular, strategies and algorithms specifically developed to process the collected data of interest will be directly implemented on edge devices, in order to produce aggregated information , reducing the total amount of transmitted information. The evaluation criterion is defined against a centralized approach taken as the baseline, with a target of at least 20% reduction in traffic transmission.
2. **KPI 4.6 (Health assessment)** focuses on evaluating the health status of a selected group of users through a defined health score derived from multiple parameters collected by wearable devices. The considered health score is the PAI, expressed as a percentage value and obtained from smartwatches insights. This KPI will be validated by comparing the computed PAIs with both subjective perceptions and objective medical diagnoses of the involved subjects.

3.7.4 User aspects: stakeholder engagement in pilot development

Within the pilot P1-TOR, different classes of stakeholders have been identified.

1. **Non-professional athletes**. This group of participants is being recruited at the scientific campus of UNIPR. The investigators are responsible for thoroughly illustrating the main objectives of the pilot, together with its potential benefits and risks. Prior to the data collection, participants are required to provide written informed consent.
2. **PD patients**. This group of participants is recruited at the “Istituto Auxologico Italiano” in Piancavallo.

3. **Medical staff.** Professionals with relevant experience in PD are involved in the pilot in order to provide guidance on the data collection process, patient management, and the interpretation of clinical outcomes.

3.7.5 User-based KPI assessment

The pilot P1-TOR adopts a user-centered approach, where aspects related to the users' experience are considered as fundamentals to evaluate the overall effectiveness of the developed health monitoring system. Accordingly, a set of user-based KPIs and their respective evaluation criteria have been identified to systematically gather insights on the participants' perceptions and experiences throughout the pilot. The identified user-based KPIs are listed in Table 19.

Table 19: user-based KPIs for pilot P1-TOR

KPI	Description	Target/Assessment Strategy
Usability rating	System usability score via post-test questionnaire	Mean score ≥ 70 (acceptable usability threshold)
Wearability score	Subjective rating of comfort and load for each sensor component	≥ 50% responses rating comfort ≥ 3 (1-5 scale)
Perceived accuracy	Subjective rating of perceived accuracy	≥ 80% responses rating accuracy ≥ 4 (1-5 scale)
Feedback usefulness	Subjective rating of perceived usefulness and understandability of the performance feedback.	≥ 80% responses rating usefulness ≥ 4 (1-5 scale)
Willingness to adopt	Likelihood to continue using the developed platform in a real-life setting and environment	≥ 60% indicate willingness to use system long-term

- **Usability Rating** refers to evaluate the system usability on the basis of: ease of use, clarity of instructions, intuitiveness of data collection interfaces, and the perceived burden of device usage.
- **Wearability score** consists in rating the comfort and unobtrusiveness of the worn sensing devices, especially in contexts of long-term monitoring and physical activity, where excessive burden or disruption are considered as critical factors.
- **Perceived accuracy** targets at measuring the degree of trust that participants attribute to the data collected by the system (e.g., sleep, physical activity, stress/load), as well as the perceived relevance in providing useful information on the health status.
- **Feedback usefulness** concerns the extent to which the performance feedback is understandable and perceived as beneficial and meaningful by the participants involved in the study.
- **Willingness to adopt** reflects the participants' attitudes towards the continued and long-term use of the developed monitoring system beyond the pilot, considering factors such as the perceived utility and burden.

The identified user-based KPIs are intended to be assessed by means of quantitative and qualitative questionnaires, which will be administered to the participants during the data collection campaign.

3.7.6 User aspects: Gender/age issues and ethical concerns in P1-TOR

At present, 13 healthy participants, recruited at the scientific campus of UNIPR, aged between 25 and 35, are involved in the pilot. All participants were asked to walk along a straight 7.5-meter path,

turning at the endpoints. The task was repeated twice: the first time at a self-pace under normal walking conditions, and the second time while simulating a Parkinsonian gait. From the collected data, key gait features—including step time, stride time and cadence—were extracted to enable a detailed analysis and highlight differences between the two walking conditions.

In parallel, 4 PD patients have been recruited at the “Istituto Auxologico Italiano” in Piancavallo. The patients were supervised by medical personnel during the protocol execution. They were asked to walk in a straight line, following a rectangular path with 90-degrees turns, wearing a single XSens Movella Dot on the lumbar region. The data acquired in Piancavallo were shared with the UNIPR research group for further analysis, including signal processing and gait feature extraction aimed at characterizing gait patterns in PD patients.

3.7.7 P1-TOR transition into cycle 2: takeaways and feedback

During Cycle 1, the pilot P1-TOR successfully demonstrated the core functionalities of the developed health monitoring system, as well as its feasibility in real-life contexts. The system validation was performed against gold-standard setups, demonstrating the reliability and consistency of data collected during the test sessions.

During Cycle 1, however, several limitations were identified that will be expedient to improvements implemented during the upcoming Cycle 2. The main challenge concerned the recruitment of a sufficient number of PD patients, which resulted in a dataset limited in size and heterogeneity. In order to overcome this limitation, parkinsonian gait patterns were simulated allowing for a preliminary analysis and system testing. Cycle 2 will focus on expanding the participant pool and seeking for a more balanced representation across all disease severity levels, thereby improving the generalizability of the results and ensuring that all classes are adequately represented.

4 UC 2 demonstrator evaluation cycle summary

Use Case 2 (UC2) will improve safety and traffic flow by offering better situational awareness to all traffic participants and automated systems.

In particular, UC2 will focus on the protection of Vulnerable Road Users (VRUs) including pedestrians, cyclists and powered two-wheeled vehicles. This will be done by combining inside vehicle observations with on vehicle and infrastructure sensors. Information will be gathered by sensors on the neighboring infrastructure, in and outside vehicles and bicycles and wearables. With the gathered sensor data, a dynamic, distributed and shared model of the situation will be built. The model will verify if the traffic participants are aware of possible dangers in the situation around them and in a state able to react.

The model is composed of digital twins that communicate with each other in a region of around 100 to maximum 200m.

UC2 will work towards this goal from two angles. The first will focus on realizing the Collaborative situational awareness at an intersection by combining observations from the various traffic participants. The second will focus on Driver distraction, where the driver's attention and possibly passengers will be monitored, and possible distraction identified. By considering the human elements in these aspects, UC2 aims to create effective and user-friendly solutions that enhance safety and efficiency in traffic scenarios, ensuring acceptance and usability among users.

UC2 consists of six demonstrators (deviating slightly from the definition of 'demonstrators' presented in section 2 of this document) that are developed in five pilots. There are two group of demonstrators with each three demonstrators.

Demo 2.1: Collaborative situational awareness to protect VRUs at an intersection

- Demo 2.1.1: Situational awareness collaboration simulation in a traffic simulator (P2-BRU).
- Demo 2.1.2: Vehicle situational awareness digital twin integrating available information from on-vehicle sensors into information that can be shared with other traffic users and capability to receive additional observations from other traffic users (P2-TMP).
- Demo 2.1.3: Demonstrator with one vehicle and a roadside unit illustrating sharing information on observed traffic situation (P2-HH).

Demo 2.2: Driver distraction detection to avoid accidents (with VRUs)

- Demo 2.2.1: Driver distraction monitoring in a driving simulator (P2-REGG)
- Demo 2.2.2: Driver and passenger mood and distraction monitoring (P2-TMP)
- Demo 2.2.3: Driver attention and traffic situation matching (P2-TMP)

As previously explained in the introduction to this document, we will polish the terminology in subsequent revisions of this document (such as D5.4) so that the glossary covers all of the different terms used in all domains and UC2 is brought more closely in line with UC1 and UC3.

4.1 P2-BRU evaluation cycle 1

The P2-BRU acronym stands for two pilots. The first one is a simulated proof of concept of using digital twins as a system of systems in Mobility. The second one aims to show, as much as possible, the combination of real hardware implementation of those concepts.

Demo 2.1.1 focuses on simulation, allowing developers to test and validate concepts in scenarios that cannot be replicated in the real world due to cost, danger, or ethical constraints. By providing a safe and controlled environment, the simulation enables thorough testing of C-ITS solutions and their potential impacts. Simulation generated by embedded processors on real hardware can complement the virtualisation of the sensors. In Demo 2.1.2 the sensors will use real detections. This will be done in cycle two.

4.1.1 Final pilot set-up

Situational awareness collaboration simulation in a traffic simulator. Vehicle situational awareness digital twin integrating available information from on-vehicle sensors into information that can be shared with other traffic users and capability to receive additional observations from other traffic users.

- Simulating traffic with cars and VRU
- Simulating RSE, V2V and V2I
- Enhanced reality: combination of real word detections and simulation.
- What if scenarios.

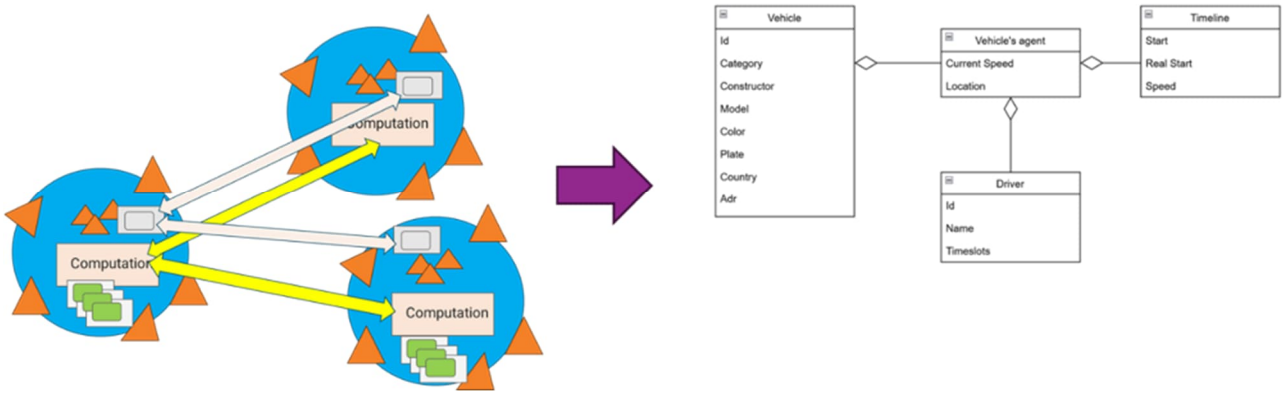


Figure 40 Digital twin concept

The Pilot is realized with the MacqSIm City Simulator. Open Street Maps (OSM) is used to create an oriented graph of the road network. Configuration files describe the roadside equipment placement of different make and models and model driver behavior patterns. The roadmap and camera placements are uploaded into the Macq Mobility Manager M3.

MacqSIm City has an agent for each vehicle and driver combination. The agent uses the oriented graph to drive from origin to destination according to its driver behavior pattern. When it passes a roadside equipment it generates detections that are sent to M3 with the corresponding communication protocol. M3 generates detections that are stored in the database and visualized in the HMI.

4.1.2 Pilot evaluation execution and protocol details

Vehicle driving simulation

When the simulation starts, all vehicles start driving somehow realistically on the road network. Each of them have a destination (a node of the road network) and computes a path from their location to that destination. They then drive on the road network according to that path. On each link of the graph (called road section) the vehicle will attribute itself a speed realistic for its environment (take into account the speed limit of the road section, the number of other vehicles on it and their average speed).

Recognitions

Each time a vehicle drives in front of a camera, a recognition is generated. All fields of the recognition are filled by MacqSim simulating the camera gathering the information from the vehicle. The recognition is then sent to M3 with the corresponding protocol. If the recognition is supposed to come from a G3, the UDP protocol is used, and if it comes from a QCAM3, QCAM5 and generic the OpenAPI protocol is used. Those recognitions can be seen on M3 and can be used by any module that only need recognition to works (trajectory control, blacklist …).

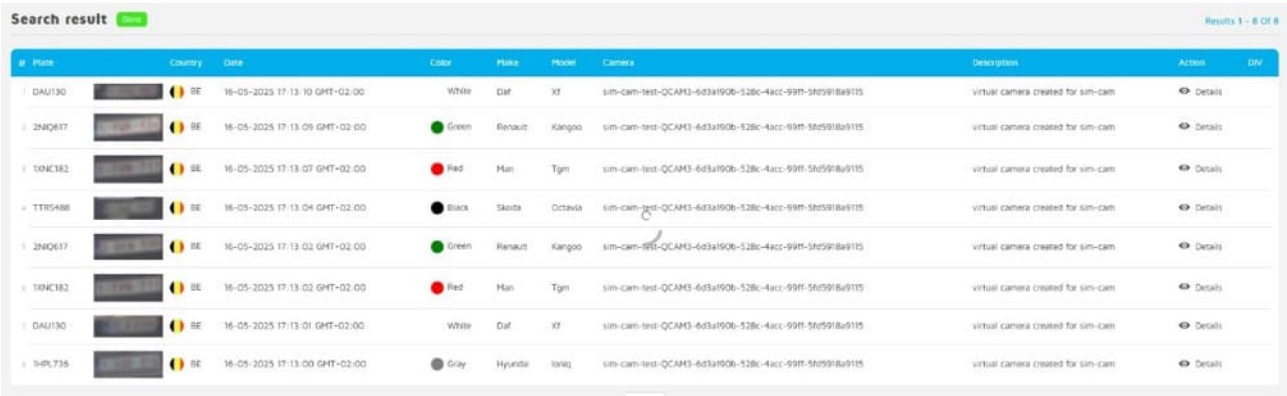


Figure 41 Detections generated by MacqSim City shown in the Macq Mobility Manager M3

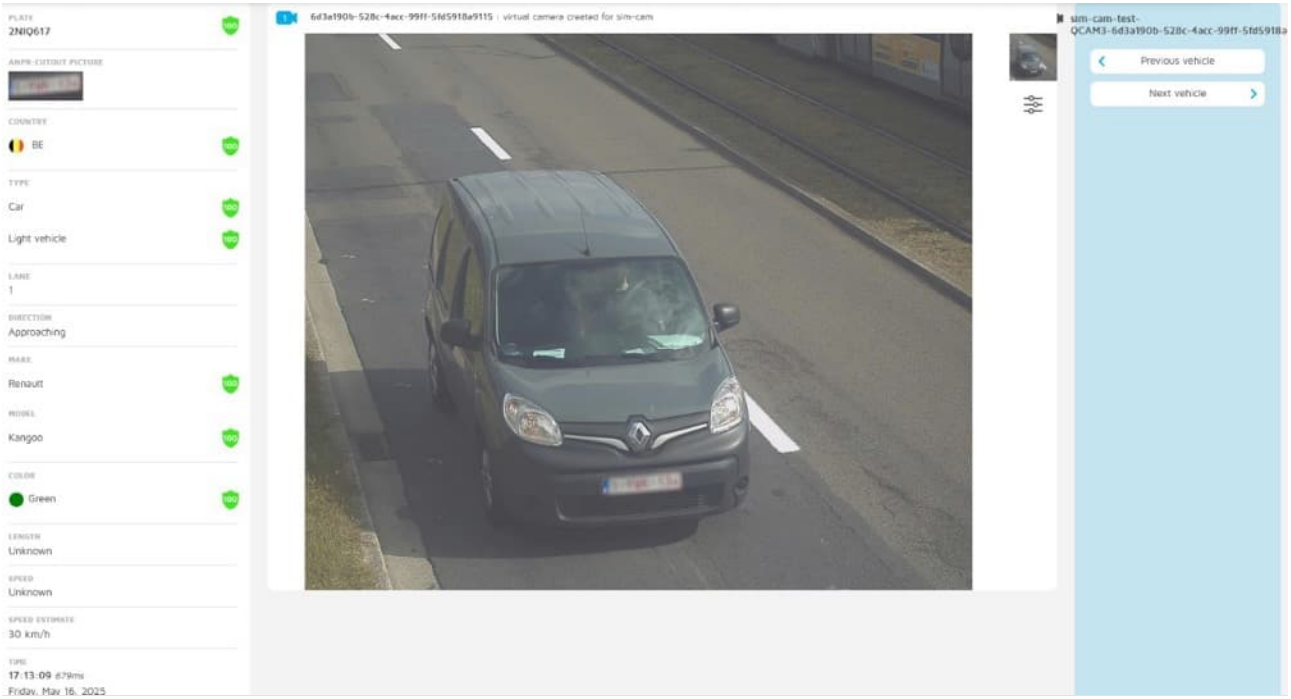


Figure 42 Details of a simulated Macq QCAM camera detection

4.1.3 Pilot technical KPI measurements

The MacqSim City simulator was tested on large cities such as Brussels and Amsterdam with 100.000 moving vehicles. Performance depends on the hardware used. On practical hardware the limit of the Macq Mobility Manager M3 handling the output are reached before the simulator. The behaviour can be real-time or faster.

4.1.4 User aspects: stakeholder engagement in pilot development

MacqSim City started as a demo program for the sales team. It evolved into a stress test system for the test automation team. Now we are turning it into a mobility simulation tool to test concept that in the current state of the art are difficult to test in real world scenarios.

4.1.5 User-based KPI assessment

Since MacqSim City is a simulation program, there a no real users involved. We hope at the end of the project to turn parts of the simulation into a real-world pilot.

4.1.6 User aspects: Gender/age issues and ethical concerns in P2-BRU

MacqSim City is a simulation tool. There is currently no simulation of age and gender aspects. There are no plans to involve the ethics focus group. This could be necessary if simulation is combined with real world data.

4.1.7 P2-BRU transition into cycle 2: takeaways and feedback

From macro to micro traffic: In the current implementation cars wait at a starting node of a segment. When enough time has elapsed according to their speed they go to the end-node of the segment at once. This is done for program optimalisation reasons and is OK if we want to simulate a complete city. For a simulation at intersection level we need the intermediate vehicle locations.

In cycle two we will the integration with the Carla and SUMO simulators.

4.2 P2-HH evaluation cycle 1

Originally, the plan for Pilot 2 in Hamburg focused on leveraging CIT's cloud infrastructure to collect and enhance sensor data from various sources. The goal was to demonstrate data fusion and real-time analysis capabilities, potentially through a V2X-DistriMuSe-Message demo. CIT was designated as the lead, as other pilots were not utilizing its developments.

4.2.1 Final pilot set-up

Following practical and legal considerations, the scope of the pilot has been refined. Instead of equipping the live operational intersection in Hamburg with a wide range of sensors from all UC2 partners (including FLIR, Xenomatix, Macq, and IMEC), we will now focus on two key components:

FLIR's thermal camera

consider it's roadside unit

This decision was driven by constraints such as power availability, mounting infrastructure (cabinets, brackets, poles), and GDPR compliance.

The revised objective is to showcase **data fusion** between sensor inputs and V2X messages. Additionally, we aim to demonstrate a **physical integration** between the FLIR camera interface board and the roadside unit from consider it, enabling **live triggers** towards the traffic controller.

This streamlined setup will allow us to maintain focus on the core goals of the pilot while ensuring feasibility and compliance.

Note: This physical integration was not part of the original project scope and introduces additional research & development effort. Due to technical complexity and resource constraints, this integration may only be realized after the official end of the project.

4.2.2 Pilot evaluation execution and protocol details

Below is an overview of the location selected for the P2-HH pilot. The first two images illustrate the site: Sandtorkai in Hamburg, situated in the heart of the city. This area is particularly interesting due to its high traffic volume and the presence of vulnerable road users, such as pedestrians and cyclists.

The road layout includes a curve and elevation changes, adding complexity to the detection and monitoring tasks. Currently, we have access to two existing poles, which will be used for sensor mounting. While the pole heights are not ideal, they are sufficient for the pilot's objectives.

Additionally, we may be able to leverage existing infrastructure already equipped with FLIR thermal sensors to supplement data collection. In the images, the red and yellow dots indicate the positions of the two poles available for use.

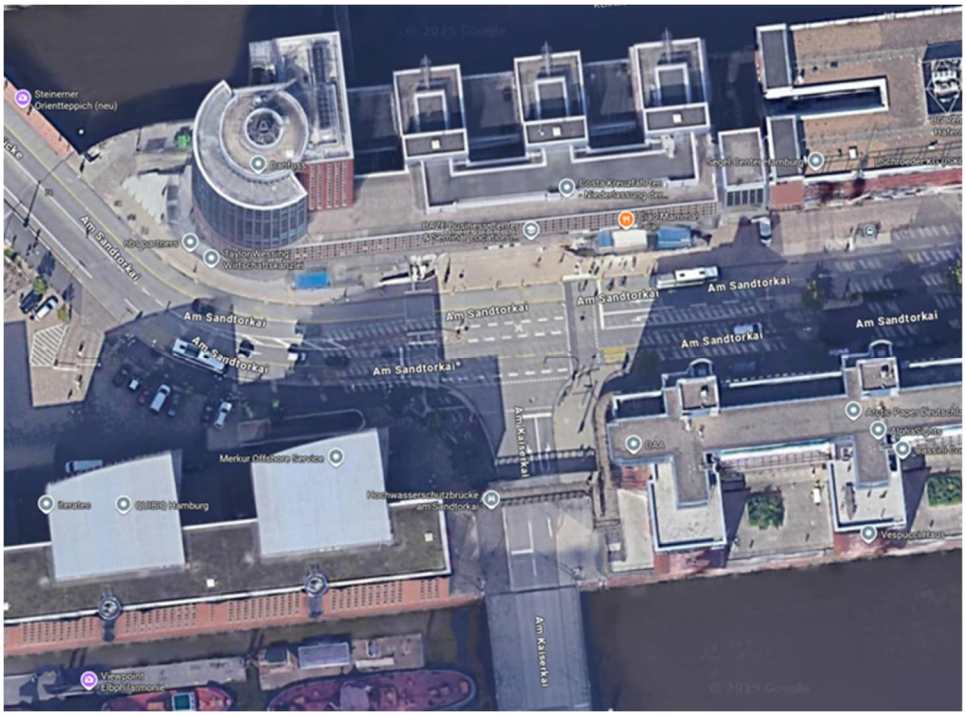


Figure 43 P2-HH chosen intersection in Hamburg

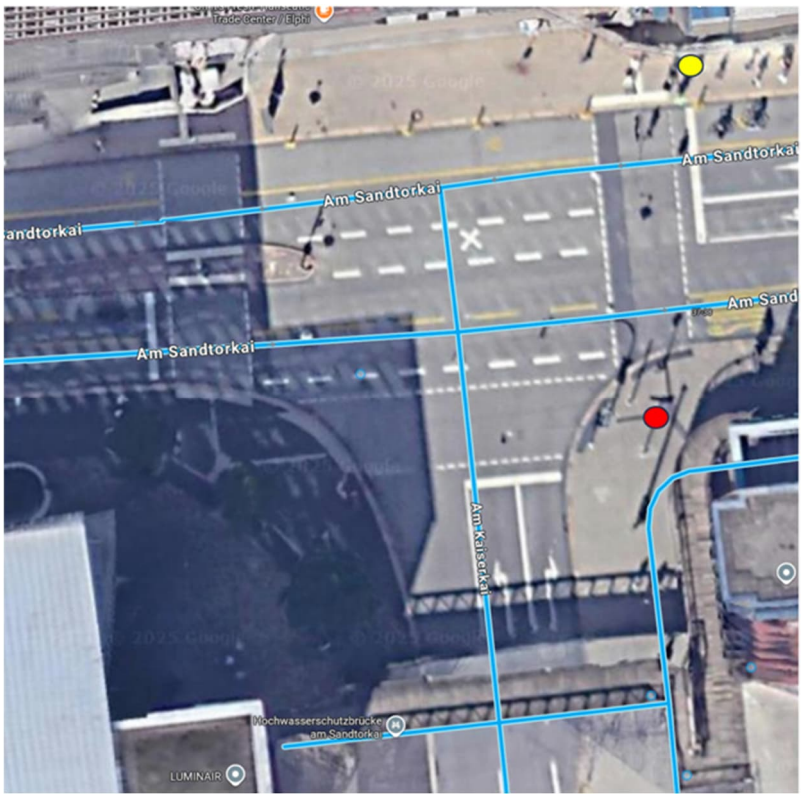


Figure 44 Detail of the used Hamburg intersection in P2-HH

Below, you'll find detailed photos of the pole and the surrounding area corresponding to the red dot location.

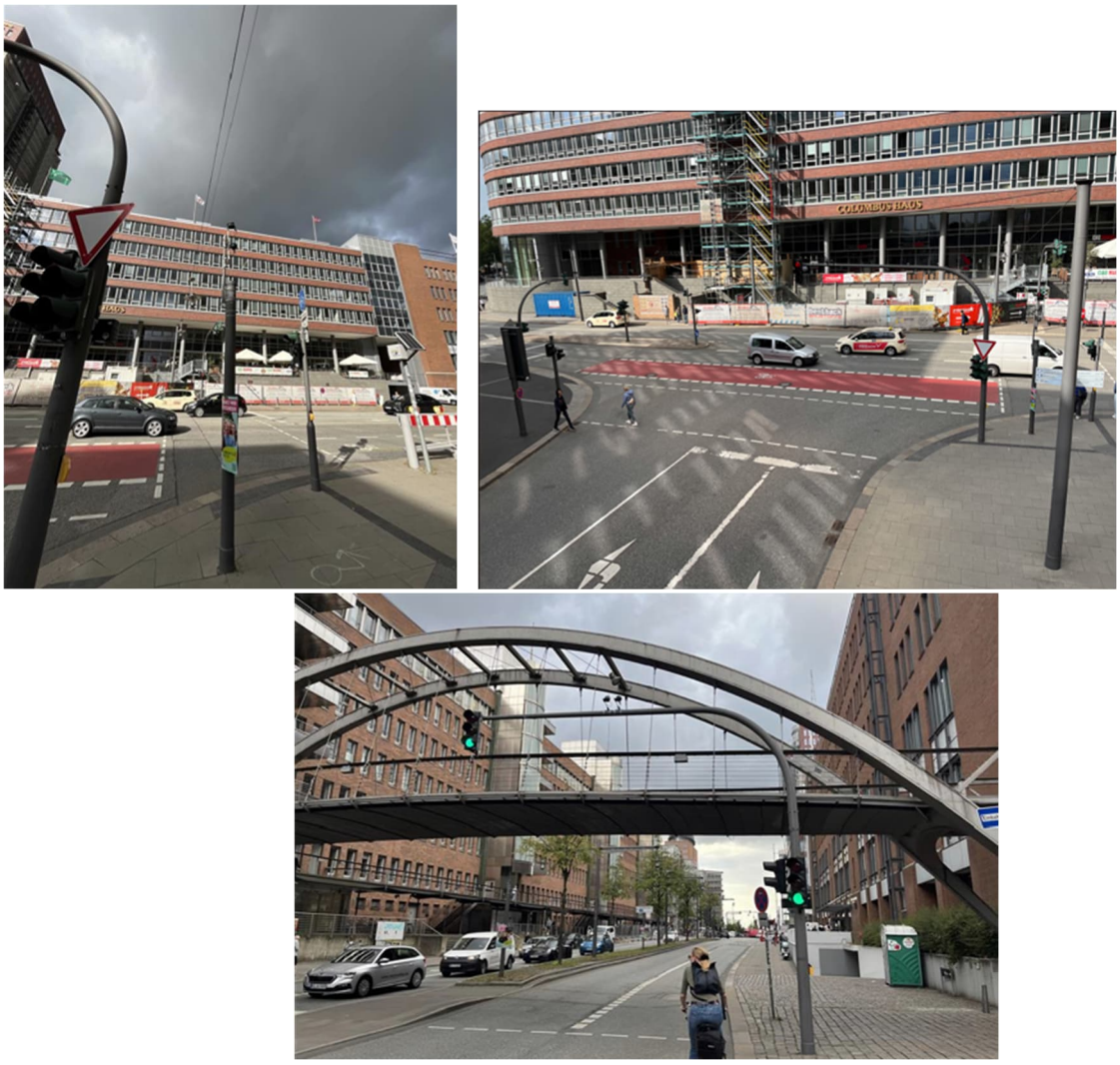


Figure 45 Ground level photos of the P2-HH target intersection

Below, you'll find detailed photos of the pole and the surrounding area corresponding to the yellow dot location.

Figure 46 Pole used for equipment mount at P2-HH

The following section outlines the planned setup of **cameras** and the **Roadside Unit (RSU)** at the pilot location.

We intend to install **multiple cameras** positioned to monitor different segments of the intersection. These will include **infrared thermal cameras**, which will be strategically placed to ensure optimal coverage. The data from these cameras will be **fused across multiple viewpoints**, enabling a comprehensive understanding of traffic dynamics and road user behavior.

A **Roadside Unit (RSU)** will be deployed to **receive all incoming V2X messages** from connected vehicles and infrastructure. The RSU will act as a central node for communication and data aggregation.

The primary goal of this setup is to **fuse sensor data and V2X messages** to identify **potential safety risks**, such as conflicts between vehicles and vulnerable road users (e.g., pedestrians and cyclists). This fusion will support real-time decision-making and contribute to safer intersection management.

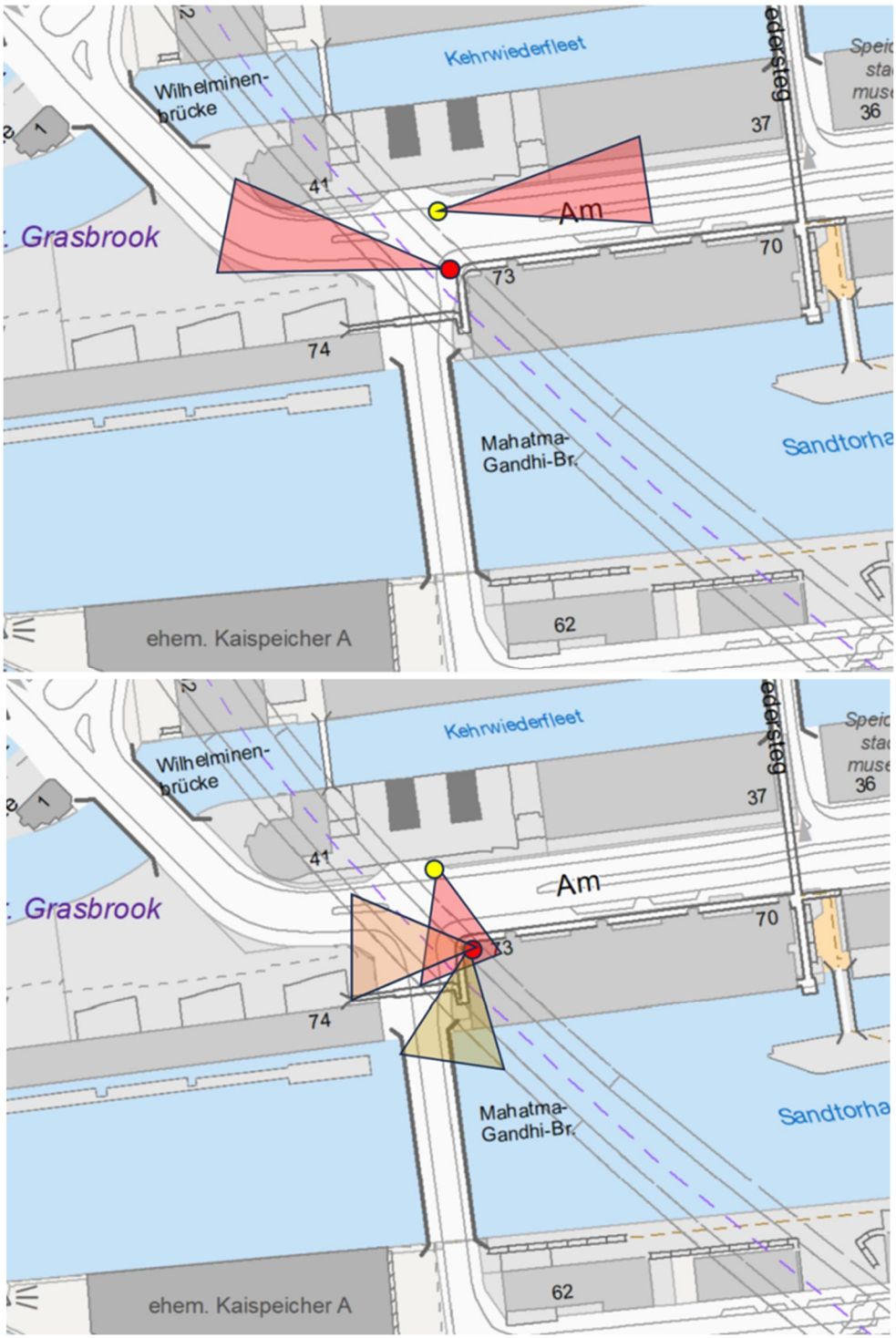


Figure 47 Summary of the used intersection in P2-HH

We follow on the description providing more details w.r.t. the used sensors:

CiT One Road Side Unit (RSU)

The CiT One RSU is a high-performance communication device designed for vehicle-to-infrastructure (V2I) applications. It is powered by a robust NXP i.MX 8XLite processor with dual 1.2 GHz cores, supported by 1 GB RAM and 8 GB internal storage, enabling efficient real-time data processing.

Security is ensured through the integrated NXP SXF1800 hardware security module, while connectivity is provided via 100 Mbps Ethernet interfaces. The RSU supports a wide range of communication standards, including:

- IEEE 802.11p for V2X
- LTE
- Wi-Fi (802.11 a/b/g/n/ac)
- Bluetooth 5.0
- GNSS via the u-blox MIA-M10Q module

All antennas are fully integrated into the housing, making the RSU compact and easy to deploy in roadside environments.

Figure 48 CiT One RSU – compact V2I communication unit with integrated antennas

FLIRs Devices

The P2-HH pilot utilizes a specialized interface designed to connect zone outputs from BPL3 sensors and/or Power Line (PL) sensors to a traffic controller. This interface provides both power and communication to the connected sensors and enables system configuration and data visualization via a PC connection.

Sensors that will be used for this pilot will be the ThermiCam AI, which is an advanced thermal imaging sensor designed for reliable detection and classification of road users in complex urban environments. The key capabilities are.24/7 thermal detection with edge-based AI for robust intersection control

Multi-object tracking in all lighting and weather conditions

Detection by class: vehicles, bicycles, pedestrians

Traffic data collection including turning movement counts and queue length monitoring

Optional wrong-way driver detection

PSH data output (Position, Speed, Heading) via API

Comprehensive reporting via Acyclica (heatmaps, movement counts)

Technical Highlights

632 variant

Detection Range: 20–122 meter for vehicle presence

Field of View: 32° horizontal × 26° vertical

690 variant

Detection Range: 0–25 meter for vehicle presence

Field of View: 32° horizontal × 26° vertical

Resolution: VGA (640×480), 30 fps

Streaming: RTSP, H.264/H.265/MPJEG

Power & Communication: PoE and Broadband over Powerline (BPL) via TI BPL3 EDGE B&SIU interface

Environmental: IP67 rated, NEMA TS2 compliant, -29 to 165°F operating range

Mechanical: Weatherproof aluminium housing with integrated sunshield

Figure 49 FLIR/CiT One RSU – compact V2I communication unit with integrated antennas

4.2.3 Pilot technical KPI measurements

consider it

The main technical KPI for the P2-HH Pilot is the use of the cloud as a method to combine or enhance data regarding traffic situations. The pilots aim to proof the feasibility of allowing multiple information flows to different communication interfaces to be combined in the cloud. After the enhancement in the cloud the improved data will be made available for further use. This will be the main KPI for the pilot. The evaluation is therefore straightforward in measuring if the current state of the art, which is collecting data from one sensor, can be improved. Improvements are counted in information flows that can be managed simultaneously.

For this pilot, we have intentionally limited the variety of sensors being deployed, as the demonstration will take place at a real-life operational intersection in Hamburg. The focus has shifted toward utilizing multiple infrared cameras from FLIR, strategically positioned to monitor different areas of the intersection.

These cameras will be fused together, combining their thermal imaging data to create a unified view of the environment. This fused sensor data will then be mapped against incoming V2X messages

received by the Roadside Unit (RSU). The goal is to integrate these data sources to identify potential safety risks, particularly those involving vulnerable road users such as pedestrians and cyclists.

In addition to the Hamburg pilot, we have defined an additional pilot in Kortrijk (P2-KORT). This second site will be used to demonstrate and evaluate the combination of different sensor technologies, providing further insights into multi-sensor fusion and its impact on traffic safety and efficiency.

The evaluation of the technical KPIs remains the same.

FLIR

The P2-HH pilot in Hamburg is designed to demonstrate the fusion of thermal imaging data from multiple FLIR infrared cameras with V2X messages received by a Roadside Unit (RSU). The setup aims to monitor a real-life intersection with high traffic complexity and vulnerable road users.

To evaluate the effectiveness of this system, we have defined the following key performance indicators (KPIs):

- **Object Detection & Tracking**: The system must detect and track road users across multiple sensors, even with minimal overlap and varying angles. Performance will be measured using the Multiple Object Tracking Accuracy (MOTA) metric, with a target of 50% improvement over current benchmarks.
- **Near Collisions:** Identify 80% of vehicles passing within 1.5 meters of vulnerable road users (VRUs).
- Achieve sub-0.5 meter localization precision using LiDAR (where applicable).
- **Detection Latency:** Ensure real-time responsiveness, with detection-to-processing latency kept below 200 milliseconds.
- **False Positive Rate:** Maintain a false detection rate below 5%, ensuring reliable identification of road users and events.
- **System Uptime:** Guarantee 99.9% operational availability, minimizing downtime and ensuring continuous monitoring.
- **Data Accuracy:** Achieve 95% accuracy in object classification and tracking across fused sensor and V2X data.

These KPIs will guide the technical validation of the pilot and help assess the system's potential for improving traffic safety and efficiency in urban environments.

4.2.4 User aspects: stakeholder engagement in pilot development

The **City of Hamburg** has been actively involved in this pilot, providing access to a real-world test location that does not interfere with operational traffic infrastructure. Their support enabled the correct setup of devices and ensured a realistic testing environment. Hamburg has also expressed interest in contributing to the **final evaluation** of the pilot by offering an additional test location and providing feedback on results and future needs.

During the pilot development, we identified the need to **integrate the CiT One Road-Side Unit (RSU)** with the existing **interface board** used for powering and communicating with FLIR's fixed-

mounted cameras. This integration aims to enable **sensor fusion**, allowing both systems to work together and provide real-time signals to the traffic controller. This would help demonstrate the **practical benefits** of the solution more effectively.

Further evaluation activities are planned to validate the scenario and refine user-based KPIs based on stakeholder feedback and technical feasibility.

4.2.5 User-based KPI assessment

At this stage, no full evaluation of user-based KPIs has been conducted for Pilot 2 Hamburg. Cycle 1 primarily involved smaller setups and individual partner tests, which were focused on technical validation rather than user experience or acceptance.

The comprehensive evaluation of user-based KPIs is planned for Cycle 2, once the full pilot setup is deployed. This will include assessments related to usability, user experience, trust, and acceptance, based on interactions with the system and feedback from target users and stakeholders.

4.2.6 User aspects: Gender/age issues and ethical concerns in P2-HH

The ethics focus group for Pilot 2 Hamburg has been formed and the session is scheduled for the end of September 2025. The results will be processed by the end of the same month. The group includes relevant stakeholders and target users, selected to ensure diverse perspectives on ethical considerations related to the pilot. The session will be conducted using the DistriMuSe Ethics Exercise Tool, as requested by WP7.

4.2.7 P2-HH transition into cycle 2: takeaways and feedback

Highlights of Cycle 1

- **Stakeholder engagement**: The City of Hamburg actively supported the pilot by providing a real-world test location that does not interfere with operational traffic infrastructure.
- **Refined pilot scope**: The pilot was adjusted to focus on two key components (cameras and RSU) instead of integrating all UC2 partner sensors.
- **Complex urban environment**: The selected location (Sandtorkai) offers high traffic volume and diverse road users, making it ideal for testing detection and monitoring capabilities.
- **Sensor fusion setup**: Deployment of multiple cameras and an RSU to enable real-time data fusion and V2X message processing for safety risk detection.

Lowlights of Cycle 1 Evaluation

- **Reduced scope**: The original plan to integrate a wide range of sensors from all UC2 partners was scaled down due to practical and legal constraints.
- **Suboptimal infrastructure**: The available poles for sensor mounting are not ideal in height, which may affect coverage and data quality.
- **Limited integration**: Full integration of all sensor systems (e.g., FLIR, Xenomatix, Macq, IMEC) was not achieved in this cycle.

Intended Changes for Cycle 2 Specification and Evaluation

- **Sensor integration improvement**: Plan to integrate the CiT One RSU with the interface board used for FLIR cameras to enable better power supply and communication.
- **Enhanced data fusion**: Research into combining both sensors to improve data fusion and provide actionable signals to the traffic controller.
- **Scenario refinement**: Adjustments to the pilot setup to better demonstrate real-world benefits and safety improvements through more effective sensor collaboration.

4.3 P2-KORT evaluation cycle 1

Pilot P2-KORT, coordinated by FLIR in collaboration with XenomatiX, IMEC, consider it, and Macq, focuses on the integration and validation of situational awareness technologies for Use Case 2, aiming to improve both traffic safety and operational efficiency.

During the Cycle 1 integration phase, it became clear that an additional pilot setup was needed. Rather than evaluating each partner's contribution in isolation, P2-KORT introduces a unified demonstrator where all relevant devices and solutions are jointly assessed. This setup enables data sharing via a common protocol, allowing for a more holistic evaluation of the integrated system and its capabilities.

The core objective of P2-KORT is to monitor traffic behaviour using a diverse array of fixed-mounted sensors installed on existing traffic infrastructure. These include visual cameras, long-wave infrared cameras, LiDAR, and radar systems, strategically positioned to provide complementary and potentially non-overlapping views. The goal is to establish a real-time situational awareness system that can detect safety-critical events and proactively alert road users to take appropriate action.

Located in the Kortrijk region of Belgium, P2-KORT forms part of the second phase of the DistriMuSe project. It complements the other pilots—P2-HH and P2-BRU—by providing the necessary infrastructure to demonstrate and validate research outcomes from multiple partners.

Unlike P2-HH, which is deployed at a live intersection, P2-KORT will be implemented as a temporary setup in a controlled environment near the FLIR office. This location was selected due to its proximity to most project partners, facilitating collaboration, practical logistics, and ensuring compliance with safety and regulatory requirements.

This pilot was conceived based on insights from the first phase of the project, where opportunities to showcase certain partner contributions were limited. P2-KORT addresses this gap by enabling the integration and demonstration of partner outputs, fostering shared valorization across the consortium.

A key innovation in P2-KORT is the development of a comprehensive situational awareness system, or "digital twin," that integrates both real and simulated data. This system will detect potential safety issues—such as near-misses or animal presence on the road—and support advanced features like object handover and re-identification, ensuring seamless tracking and management.

Additionally, the pilot will demonstrate real-time multi-camera and multi-sensor fusion tracking, showcasing distributed capabilities that allow for semi-automatic performance enhancement over time.

4.3.1 Final pilot set-up

As location we will investigate if the FLIR office in Marke nearby Kortrijk would be a good candidate to illustrate the different devices for UC2. We expect to have several iterations where the partners will visit the location make temporal setup and try showcase to the other partners what was achieved so far. We will define where each partners devices can be placed what the limitations we do have and how we can combine the devices & data in the best way to achieve a fully situational awareness. Below we will discuss each partners device and setup we already have define the common protocol to share the information which each other so a fusion component can be created and shared with the roadside unit of consider it that will be able to communicate with the vehicles that have an onboard unit

The pilot setup for **P2-KORT** will be established in a **controlled environment** near the **FLIR office in Marke** (cfr. photos below), close to **Kortrijk, Belgium**. This location is under consideration due to its proximity to the majority of project partners, which facilitates frequent collaboration, practical logistics, and iterative testing.

Rather than deploying the system at a live intersection, the setup will be **temporary and modular**, allowing partners to install, test, and refine their devices across multiple iterations. These sessions will enable each partner to showcase their progress, validate integration steps, and jointly evaluate system performance.

Key aspects of the setup include:

Device Placement Planning: We will define optimal locations for each partner's devices, considering field of view, environmental constraints, and potential interference. This includes identifying limitations such as mounting options, power availability, and connectivity.
Iterative Integration: Partners will visit the site in scheduled phases to install and test their components. This iterative approach supports continuous improvement and collaborative troubleshooting.
Data Sharing & Fusion: All devices will communicate using a **common protocol**, enabling seamless data exchange. This shared data will feed into a **fusion component**, which aggregates sensor inputs to build a unified situational awareness model.
Roadside Communication: The fused data will be transmitted to the **roadside unit (RSU)** provided by **consider it**, which is capable of communicating with vehicles equipped with **onboard units (OBUs)**. This enables real-time alerts and interaction with road users.
Use Case 2 Focus: The setup is specifically designed to support **Use Case 2**, emphasizing safety and efficiency improvements through multi-sensor integration and real-time awareness.

In the following sections, we will detail each partner's device, its role in the setup, and how it contributes to the overall system. We will also outline the agreed-upon protocol for data exchange and the architecture of the fusion component.

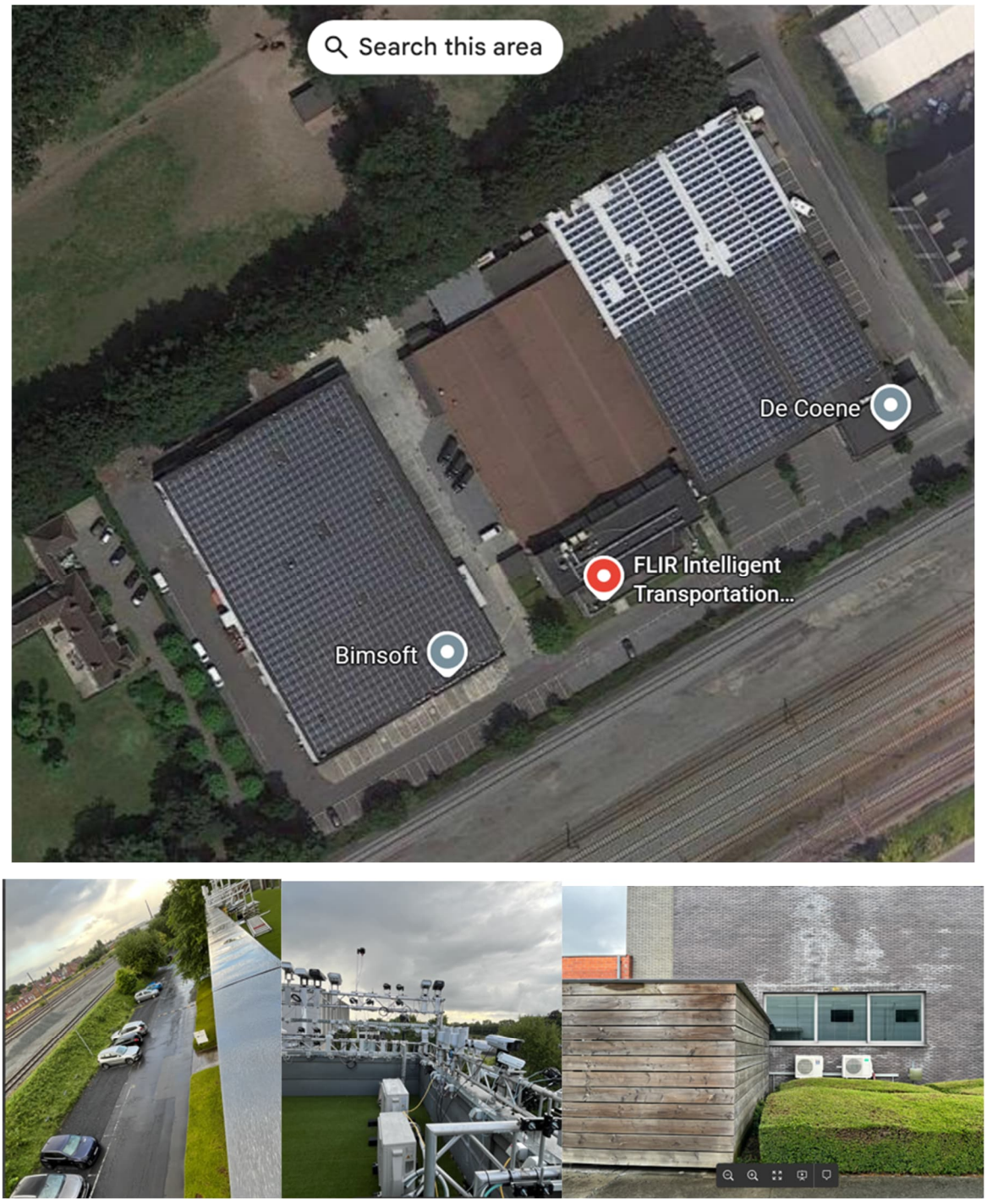


Figure 50 P2-KORT deployment area images

Objective of the XenomatiX Demonstrator in the DistriMuSe Pilot

The goal of XenomatiX within this pilot is to demonstrate real-time detection of road users using a roadside 360-degree LiDAR system. The detected objects will be shared with project partners, who will focus on fusing this data with other sources and visualizing it on a map.

The main objective of the demonstrator is to showcase the added value of combining LiDAR and camera technologies:

Camera systems offer high-resolution imagery, making them particularly effective for detecting objects at long distances.

LiDAR systems, on the other hand, excel in low-light conditions and provide more accurate distance measurements.

By integrating both technologies, the demonstrator aims to highlight how their complementary strengths can lead to more robust and reliable road user detection.

In **Figure X**, we illustrate the current equipment and setup used in the pilot. The image shows, from left to right: the **GPC360** (360° LiDAR), the **Xavia 6D** (solid-state LiDAR with integrated camera), and an example of the complete setup including a **heavy-duty tripod**, **battery**, and **router**.

Figure 51 Overview of the equipment used in the pilot from XenomatiX.

From left to right: the **GPC360** (360° LiDAR), the **Xavia 6D** (solid-state LiDAR combined with a camera), and an example of the full setup including a **heavy-duty tripod**, **battery**, and **router.**

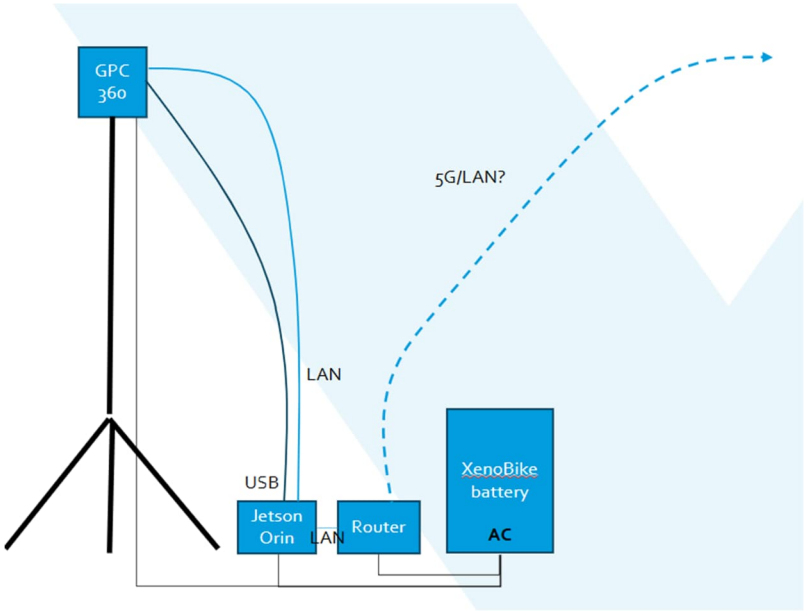


Figure 52 Setup of the Xenomatix elements in P2-KORT

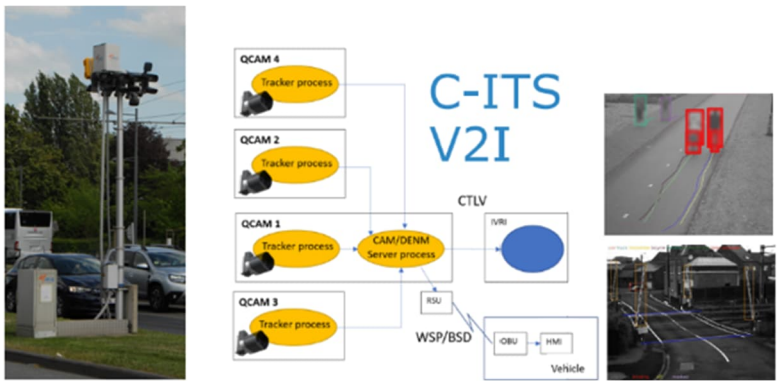


Figure 53 P2-KORT consider-it setup

4.3.1.1 IMEC

IMEC is advancing situational awareness in the P2-KORT pilot project by developing sophisticated information fusion techniques that integrate data from roadside infrastructure and vehicles. This data is transmitted via V2X communication to enhance real-time detection, tracking, and the continuous updating of a digital twin that mirrors the traffic environment.

To achieve this, IMEC is deploying sensor boxes equipped with RGB cameras, thermal imaging, radar, and LiDAR. These sensors enable both feature-level and decision-level data exchange between infrastructure and vehicles. The result is a system capable of aggregated detection and tracking, dynamic updates to the digital twin, and context-aware classification powered by AI models running at the edge.

Figure 54 Imec sensor boxes with visible and thermal camera, radar and lidar

Currently, imec utilizes two sensor boxes, each capturing data from the following sensors: FLIR TrafiSense Dual 690 thermal/visible camera, TI AWR 1443 radar and Xenomatix Xact or Livox Horizon lidar. These sensor boxes communicate via LAN network, exchanging high-level decisions or features, in order to improve the detection and prediction accuracy.

For example, Figure 55 shows an example of a car turning right at the intersection. The car is being tracked by the first sensor box. Once it arrives close to the coverage zone of the second sensor box, the first sensor box issues a warning message to the second sensor box, that the vehicle is arriving. After receiving this message, the second sensor box can boost the confidence of detection and track the car further.

Figure 55 Collaborative sensor fusion with two sensor boxes containing multiple sensors

As part of the pilot, IMEC is focusing on several key developments. These include the design and calibration of sensor boxes with embedded low-level processing, the creation of detection, tracking, and prediction algorithms based on fused sensor data, and the integration of vehicle data into the broader detection and classification pipeline. Additionally, IMEC is working on AI models that can adapt in real time to changing traffic conditions, supported by distributed learning frameworks that allow for scalable deployment.

Through P2-KORT, IMEC aims to demonstrate how the fusion of infrastructure and vehicle data can significantly improve the accuracy and timeliness of situational awareness, ultimately contributing to safer and more efficient traffic systems.

4.3.1.2 FLIR

As part of the P2-KORT pilot, FLIR is preparing a flexible and comprehensive camera deployment around its office premises. The setup will include installations on the building's infrastructure, as well as temporary configurations using two available trailers. This dual approach allows FLIR to strategically position devices for optimal coverage and data collection.

Given that the pilot operates within a controlled environment, FLIR will leverage its full range of devices, including those at lower Technology Readiness Levels (TRL). This presents a unique opportunity to test and validate emerging technologies, such as the interface module designed to fuse data from multiple sensors.

The devices planned for deployment include:

TrafiCam AI – visual cameras for detailed scene analysis
ThermiCam AI – thermal cameras for enhanced detection in low-visibility conditions
TrafiBot Dual AI – dual-sensor units combining visual and thermal capabilities

These devices will work together to provide a rich, multi-modal data stream, supporting advanced fusion and situational awareness. The pilot will serve as a testbed for FLIR's integrated sensing technologies, helping to refine detection, classification, and tracking capabilities in real-world traffic scenarios.

Figure 56 FLIR setup for P2-KORT

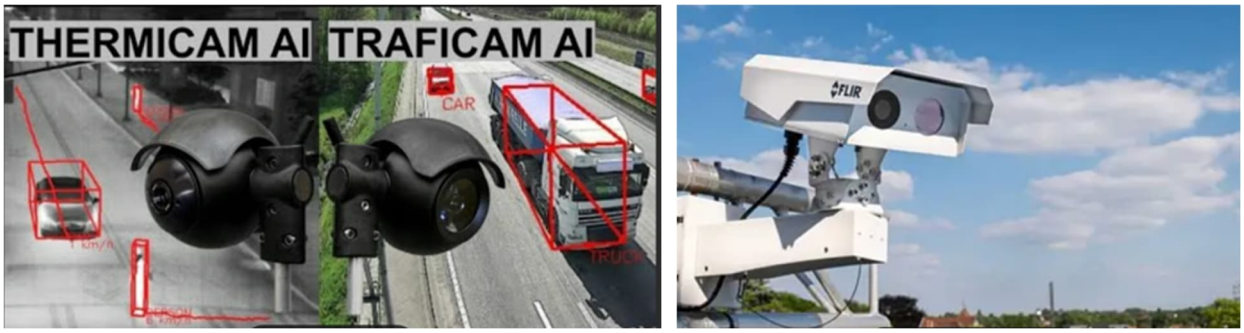


Figure 57 FLIR devices used in P2-KORT

consider it

See section 4.2 P2-HH evaluation cycle 1 since the setup and devices are identical.

4.3.2 Pilot evaluation execution and protocol details

The P2-KORT pilot is part of the second phase of the DistriMuSe project and will be executed in a controlled environment near the FLIR office in Marke, Belgium (cfr. Setup above). Unlike other pilots (P2-HH), P2-KORT is designed as a modular and temporary setup that allows partners to install and test their technologies iteratively. This location was chosen for its proximity to most partners, enabling frequent collaboration and practical logistics.

The pilot introduces a unified demonstrator where all partner technologies are jointly evaluated. Instead of isolated testing, devices share data via a common protocol, allowing for integrated system assessment. The setup includes visual cameras, thermal imaging, radar, and LiDAR sensors mounted on buildings and trailers, offering complementary views of the traffic environment.

Each partner will define optimal device placement, considering environmental constraints and connectivity. Data will be shared using the agreed protocol and fused into a unified model. The fused data will be transmitted to the RSU, which communicates with vehicles equipped with onboard units.

The pilot supports Use Case 2, focusing on improving traffic safety and efficiency through multi-sensor integration and real-time awareness.

4.3.3 Pilot technical KPI measurements

Methodological Adjustments

During Cycle 1, we identified that some of the originally proposed metrics—particularly the **Multiple Object Tracking Accuracy (MOTA)**—did not fully align with the practical setup and goals of the P2-KORT pilot. While MOTA was initially selected to measure improvements in object detection and tracking (with a target of 50% improvement), it became evident that this metric does not adequately reflect the spatial accuracy required in our context.

As a result, we are exploring alternative metrics that are more **location-aware**, focusing on **real-world positioning on a map** rather than abstract tracking accuracy. This shift better supports the integration of multi-sensor data and reflects the operational needs of the pilot.

Evaluation of Technical KPIs (

The following KPIs are being evaluated, with adjustments based on Cycle 1 learnings:

Object Detection & Tracking
Goal: Detect and track objects across different sensors with minimal overlap and varying angles.
Status: Metric under revision; moving toward map-based spatial accuracy.

Near Miss Detection
Goal: Identify 80% of vehicles passing within 1.5 meters of Vulnerable Road Users (VRUs).
Approach: Long-term data collection at a nearby site to illustrate this KPI, as P2-KORT focuses on integration rather than direct deployment.

Detection Latency
Will be evaluated within the P2-KORT setup, focusing on system responsiveness across integrated components.

False Positive Rate
Measurement planned at a separate location with long-term deployment. Evaluation will begin toward the end of Cycle 2 once the system reaches sufficient maturity.

Additional KPIs for Cycle 2

Based on lessons learned, we are expanding our KPI focus to include:

System Integration Quality
Assessing how well the different partner technologies (Xeno**m**atiX, IMEC, MACQ, **c**onsider **i**t, FLIR) can be fused into a coherent solution.

Power Consumption
Evaluating the energy efficiency of the integrated system, which is critical for real-world deployment.

Individual vs. Combined Performance
Measuring how each system performs independently and what improvements in VRU safety can be achieved through their combination.

Additionally, it's important to note that for FLIR and consider It, the P2-HH pilot plays a more central role in illustrating their specific KPIs, given its real-world deployment context. In contrast,

P2-KORT will primarily serve as a testbed for integrating and combining the outputs from the various sensor systems. This setup allows us to evaluate how the individual technologies complement each other and what added value their fusion can bring in terms of technical performance and VRU safety improvements.

4.3.4 User aspects: stakeholder engagement in pilot development

For pilot P2-KORT, the focus in Cycle 1 has primarily been on internal evaluation of the research outcomes and technical integration of partner contributions. Unlike P2-HH, where the City of Hamburg was actively involved due to the real-world deployment at a traffic intersection, P2-KORT has not yet engaged external stakeholders or end users in the same way.

At this stage, our priority has been to establish a robust setup that enables the integration and fusion of the various technological components provided by the partners (XenomatiX, IMEC, MACQ, consider It, and FLIR). This foundational work is essential before meaningful stakeholder engagement can take place.

However, we recognize the importance of involving stakeholders and target users to validate scenarios and refine user-based KPIs. Therefore, once the integrated setup is operational, we plan to:

- Engage relevant stakeholders through consultations or workshops to gather expectations and feedback.
- Involve target users (e.g., mobility planners, traffic safety analysts, or city representatives) in evaluation activities such as demonstrations or observational testing.
- Use this input to validate pilot scenarios and refine KPIs based on real-world relevance and usability.

This engagement is planned for Cycle 2 and will be critical to ensure that the pilot outcomes are aligned with user needs and can be effectively valorised post-project.

4.3.5 User-based KPI assessment

At this stage, no full evaluation of user-based KPIs has been conducted for P2-KORT. Cycle 1 primarily involved smaller setups and individual partner tests, which were focused on technical validation rather than user experience or acceptance.

The comprehensive evaluation of user-based KPIs is planned for Cycle 2, once the full pilot setup is deployed. This will include assessments related to usability, user experience, trust, and acceptance, based on interactions with the system and feedback from target users and stakeholders.

4.3.6 User aspects: Gender/age issues and ethical concerns in P2-KORT

The ethics focus group for P2-KORT (which will be held together with the P2-HH) has been **formed** and the session is **scheduled for the end of September 2025**. The results will be processed by the end of the same month. The group includes relevant stakeholders and target users, selected to ensure diverse perspectives on ethical considerations related to the pilot. The session will be conducted using the **DistriMuSe Ethics Exercise Tool**, as requested by WP7.

4.3.7 P2-KORT transition into cycle 2: takeaways and feedback

Highlights of Cycle 1

Successful engagement and collaboration across multiple partners.
Initial technical validation of individual partner contributions.
Valuable insights into the complexity of integrating multi-partner outputs.

Lowlights of Cycle 1 Evaluation

Demonstrating the combined output from all partners at the P2-HH was not feasible due to practical and legal constraint (cfr. P2-HH).
Limited opportunity to showcase interoperability and added value of combined partner technologies within the original timeline.

Intended Changes for Cycle 2 Specification and Evaluation

As a key lesson learned, we have decided to initiate an additional pilot, which was not originally planned at the start of the project (this P2-KORT).
This pilot aims to foster closer collaboration between all relevant partners: XenomatiX, IMEC, MACQ, Consider It, and FLIR.
The goal is to create a setup that not only illustrates each partner's individual contribution but also explores:

- How these technologies can be integrated.
- What added value each partner brings to the combined solution.
- How this integration can support post-project valorization.

This approach will allow us to better evaluate the synergies between partners and ensure that the final outcomes are more representative, impactful, and aligned with real-world deployment scenarios.

4.4 P2-REGG evaluation cycle 1

4.4.1 Final pilot set-up

As anticipated in Deliverable 5.1 and Deliverable 5.2, an advanced driving simulator was used to perform the tests. The following paragraphs describe the hardware, software and simulated vehicle configurations details of the final pilot set-up.

Hardware Configuration

Figure 58 shows the simulation cabin, where the control peripherals - pedals and gear lever - transmit data to an acquisition board via USB connection. The steering wheel also communicates via USB through its base, which integrates all inputs from the connected peripherals (pedals and gear). Additional driving functions, such as turn indicators, light controls, and acoustic warnings, are simulated using the programmable buttons integrated in the steering wheel.

The acquisition PC, located in the control room, communicates with the simulator PC which processes the peripheral data and manages the simulation graphics through a dedicated GPU. The simulator's visual output is distributed across three main monitors, while a dedicated HMI display shows dashboard information during driving.

The system also includes additional peripherals such as audio interface, webcam and microphone, as well as Ethernet adapters for network communication. To the right of the driver, a tablet is positioned to serve as a distraction element.

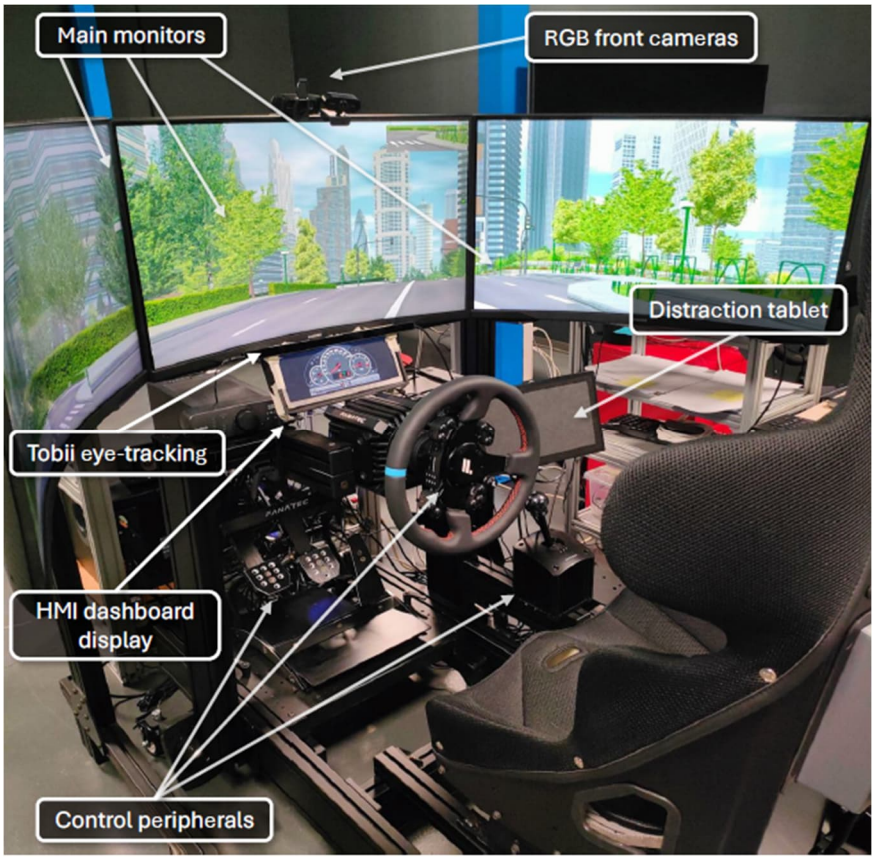


Figure 58 Hardware configuration of pilot 2.2.1

Experimental Setup

The experimental setup has undergone a few updates compared to the initial proposal and the following sensors have been integrated in the structure:

(1) Lateral camera to record entire driver figure
(2) Wearable device for heart rate measurement
(3) Eye tracking for gaze detection (experimental)
(4) Second camera in front of the driver for emotions and distractions recognition

The final configuration includes multiple data acquisition systems working in synchronization. A **frontal camera** captures frames of the user's face, synchronized with frames acquired from the simulator's visuals through matching frame numbers and timestamps. A **lateral camera (1)** provides complete coverage of the driver and their posture throughout the driving simulation session. This second camera was added to conduct further analysis on the driver, which will remain within the scope of the simulation: posture analysis through skeletal landmark recognition using Deep Learning and Computer Vision software. The aim in this case would be to approach a further level of analysis to assess any correlations between driving status and states of discomfort. Although this additional

camera has been added for filming, posture analysis will not be addressed before the second cycle. Physiological monitoring is achieved through a **wearable device (2)** the Coospo HW9 that is a wrist-worn optical heart rate monitor designed for continuous acquisition of physiological signals during physical activity. The device measures heart rate (HR) and, when signal quality permits, RR intervals, which can be used for heart rate variability (HRV) analysis and arousal estimation. Data are transmitted in real time via Bluetooth Low Energy (BLE) and ANT+ (Adaptive Network Topology) protocols, enabling integration with external applications and platforms. Visual behaviour analysis is conducted using a Tobii **eye-tracker (3)**, enabling the study of users' visual behaviour by analysing where they look and for how long, providing insights into attention patterns and visual focus distribution. An **additional frontal camera (4)** connected to a dedicated computer performs real-time emotion detection and distraction recognition by analyzing user actions. This multi-camera setup ensures comprehensive monitoring of both physiological and behavioral responses during the simulation.

Data Collection and Communication

Data integration is managed through an **MQTT broker**, which synchronizes streams from different sources, including vehicle dynamics, physiological signals, and behavioral data. Visual data is recorded with **OBS Studio** and streamed in real time via **RTMP protocol** to AITEK servers for synchronized storage and analysis. Critical events during the simulation are logged through Python scripts leveraging SCANeR APIs, ensuring traceability of key occurrences.

Software Configuration

To develop the driving scenario logic, collect data and perform simulation tests, we use SCANeR Studio software (by AVSimulation), version 2023.4. SCANeR Studio is a comprehensive software suite specifically developed for simulation in the automotive and transportation sector, particularly suitable for vehicle system development and testing. It provides a detailed virtual environment for simulating driving scenarios, traffic, sensors and human-machine interactions, offering realistic tools for creating and managing virtual environments including road networks, traffic patterns, weather conditions and lighting scenarios.

The software distinguishes itself through its modularity and flexibility, allowing users to adapt the platform to their specific needs. It provides an open architecture that enables integration with other tools and systems, making it suitable for a wide range of automotive simulation applications. Beyond its standard modules for managing visual, audio, traffic and physics entities, the platform allows for the creation of custom modules, which we have implemented for image and data acquisition.

For conducting these tests, we utilized the 3D maps provided by the software as a foundation, programming and managing all event logic while simultaneously tracking every event and situation that occurred during the simulation sessions. The system offers powerful tools for analysing telemetry and events recorded during driving sessions, enabling comprehensive evaluation of vehicle behaviour and human-machine interaction under various simulated conditions.

Simulated Vehicle Configuration

Within the software, we have configured a Supermini-class vehicle with an internal combustion engine and automatic transmission. The navigation system provides route information through directional indicators (arrows), while the dashboard displays speed, engine RPM and gear engaged. Speed limit warnings are displayed during driving. The side and central rearview mirrors have been

simulated and distributed across the three main monitors to ensure complete visibility of the driving environment.

4.4.2 Pilot evaluation execution and protocol details

This paragraph is derived from Deliverable 5.1 (Chapter 3.4 P2-REGG/ Demo 2.2.1 specification and planning, paragraph "Pilot components and proof of concepts ") and provides information about pilot architecture and the technologies that are going to be tested in Pilot 2.2.1.

The high-level architecture of Pilot 2.2.1 consists of both real and simulated components that can monitor driver behavior and collect data useful for its prediction. The system is built around the digital twins and a Decision Support System (DSS) that, receiving input data from these digital twins, will enhance driver and vehicle safety. RGB cameras track head pose, gaze direction, and behavioral patterns (e.g., mobile phone use, eating), while AI-powered emotion detection evaluates valence, engagement, and emotional state using CNN models.

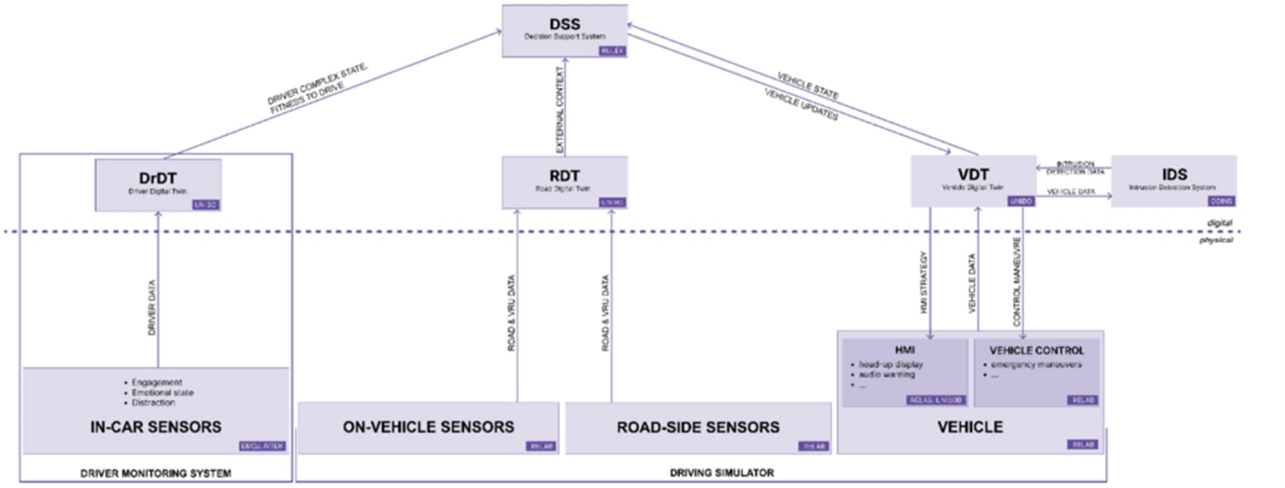


Figure 59 P2 – REGG Architecture: the figure shows the main components of the proposed system

The Driving Simulator Simulates realistic vehicle controls, ADAS features, and autonomous driving scenarios. It Integrates sensor-based capabilities (e.g., LiDAR, RADAR, GPS, V2X communication) for real-world testing conditions and can Model complex interactions with Vulnerable Road Users (VRUs) (pedestrians, cyclists) to enhance safety features. The driving simulator also simulates traffic conditions, road layouts, environmental factors, and infrastructure communication (V2I). It can test vehicle responses in various traffic and weather conditions to optimize control strategies.

By validating these Proof of Concepts, Pilot 2.2.1 aims to refine driver monitoring systems, enhance vehicle safety, and develop innovative AI-driven solutions for accident prevention in complex road environments.

Simulator Infrastructure for the Experiment: integration of pilot execution compared to D5.1

The technological infrastructure supporting the experiment is based on a distributed architecture designed to optimize the use of available computational resources. The system relies on two main workstations, each dedicated to specific computational tasks according to its hardware capabilities.

The workstation in the simulation room represents the computational core of the system. Equipped with an NVIDIA RTX 4080 GPU, 64 GB of RAM, and an AMD Ryzen 9 processor, this high-performance configuration is essential for managing the most demanding processes of the experiment. This machine handles the real-time rendering of 3D graphics on the simulator's three monitors and digital dashboard, ensuring constant frame rate and minimal latency for a smooth and realistic driving experience. It also processes all input signals from steering wheel, pedals, and other controls. In addition, the workstation manages video streaming via OBS Studio and records data from the eye-tracking system—tasks that require significant memory resources and parallel processing capabilities.

The control room workstation, although less powerful in terms of hardware, plays an equally critical role in experiment management. It runs the SCANeR Studio simulation software and coordinates modules that do not require intensive GPU acceleration. These include the **Sound module**, which manages the simulation's audio environment; the **ModelHandler**, responsible for loading and handling 3D scene elements; the **TrafficDriver**, which controls virtual traffic behaviour through AI-based algorithms; and the **Scenario module**, orchestrating the sequence of simulation events. The **Record module** captures structured telemetry data via the integrated Analysis Tool, while the **ImageSharing module** enables the distribution of visual data across components. This workstation also hosts the Mosquitto MQTT broker, which acts as the backbone of real-time inter-process communication and data collection.

OBS Studio is configured to acquire the live video feed from the simulator's three monitors. This continuous video stream is transmitted to the project partner AITEK using the **RTMP (Real-Time Messaging Protocol)**, enabling stable low-latency streaming and synchronized storage of all experimental data. To optimize bandwidth while preserving sufficient quality for analysis, the video is compressed to 640×640 resolution. A visible timestamp overlay embedded in the video guarantees frame-accurate synchronization with other data streams.

MQTT Communication Architecture

The Mosquitto broker, running on the control room workstation, provides the core of the experiment's real-time communication system. Through the lightweight and efficient MQTT protocol, multiple critical data streams are published and collected for driver behaviour analysis. Vehicle dynamics are monitored by a custom C++ module that leverages SCANeR Studio's native APIs, particularly the *VehicleDynamics* interface. This module acquires a rich set of vehicle parameters at a sampling rate of 2 Hz, enabling detailed monitoring of driving behaviour.

4.4.3 Pilot technical KPI measurements

As described in Deliverable 5.1, the following technical KPIs have been identified. No changes in methodology to assess the KPI have been made.

Table 18. indicates technical KPIs that were assessed in pilot P2-REGG

Nr	State of the art	Innovation beyond SotA	KPI	Start state	Target	Cycle 1 validation
3.6	**Federated learning-based in-vehicle Intrusion Detection System (IDS)** – Deep Learning (DL) models are commonly employed as the base models for federated learning-based in-vehicle IDSs[9], [10]. While these models are capable of detecting a wider range of network anomalies than classical ML models, in-vehicle systems are inherently resource-constrained and may not possess the required resources to accommodate sophisticated DL models. Furthermore, the existing IDSs mostly rely on analyzing network traffic, without considering the behavioural data of the embedded devices (e.g., ECU and sensor modules), typically leading to low generalization and accuracy.	We intend to enhance the level of security in edge vehicular networks through a novel in-vehicle anomaly-based Intrusion Detection System (VAIDS), based on Federated Learning that will take into consideration abnormalities not only in data flowing through the in-vehicle CAN bus (Controller Area Network) but also in behavioural data of devices (e.g., sensors) of the in-vehicle system. The developed VAIDS will take advantage of the Federated Learning technology not only to detect intrusions with low false positive rate and false negative rate but also to maintain data privacy by performing local training and inference of detection models, using TinyML[11], while at the same time reducing resource utilization of the centralized Edge Server (i.e., Federated Server).	The implemented decentralized federated learning-based in-vehicle IDS will be evaluated based on its i) time cost per execution of client selection, (ii) training time cost, (iii) accuracy, (iv) precision, (v) recall, and (vi) f1-score	i) Time cost per execution of client selection: 3ms; ii) Training time cost: 12 s; iii) Accuracy: 86%; iv) Precision: 89%; v) Recall: 77%; vi) f1-score: 86%	i) 7% improvement; ii) 7% improvement; iii)7% improvement; iv)7% improvement	i) 2.5ms ii) It depends on the hardware resources and number of clients. With 40 clients simulated simultaneously only in an AMD Ryzen 7 CPU without GPU and 32 GB RAM, a 35 round training lasts for about 30 minutes. iii) Accuracy: 88%. iv) Precision: 0.8982925. v) Recall: 0.9036383. vi) F1-score: 0.8998342.
4.2.1	**Mental state extraction** – Relies on the reliability and quality of the psychophysiological and behavioural signals recorded during a period of interest. In real life context, mental state detection relies on heart rate dynamics and in some cases, skin conductivity variation. Also, micromovements, gestures, group dynamics (positions and distances of persons in a group) and facial expression analysis can be used to complement the measurements.	We aim to improve understanding of the relationship between the various signals and subjective mental state.	Improvement in the accuracy of mental state recognition in a lab environment	Arousal and valence detection with approx. 85% accuracy	Accuracy of >90%	To be evaluated
1.4	**Camera-based sensing,** particularly 3D imaging, is used to create a true-to-life 3D presentation of the surroundings. This is used to detect the positions of humans,	We will improve the robustness and speed of the camera-based sensors to generate fewer false observations. We will make the operation robust	Computing latency Frame rate detection and mapping accuracy	>10 ms 30 fps Do not work in both night and day	Accuracy detection and mapping >90%	<10 ms in embedded GPU 100 fps

	vehicles and objects in close proximity – The ambient illuminance affects the performance (high ambient illuminance may affect the precision, and low levels the quality of color data). Existing algorithms are too slow and yield too many false detections**.**	to varying daylight conditions (e.g., darkness and sunlight).				
4.7	**Face-based driver monitoring** – There exist systems for face-based monitoring[6] but there are limitations, for example, it is difficult to analyze face material from challenging camera angles or in a low-light scenario.	We intend to improve the ability of AI to analyze faces visually in various circumstances, which may involve challenges such as low light, difficult camera angles, or previously non-detectable facial expressions.	Maximum yaw (horizontal displacement) of the face, still allowing the system to extract facial features so that the output of the extraction is adequate for further AI-based processing steps	Up to 30 degrees of yaw of the face allowed	Up to 40 degrees of yaw of the face allowed	To be evaluated
2.6	**Explainable AI** – Current AI systems should provide clear and understandable explanations of their decision-making processes to humans. However, current algorithms provide limited interpretability restricted to simple abstracted models, curbing its broader usage.	Development of new explainable AI techniques and tools that can handle complex and heterogeneous data (e.g., images or time-series) as well as complex neural network architectures (e.g., Transformers). These results are highly valuable to assess the efficacy and safety of AI-based solutions for practical applications in health, robotics and mobility. Design of a feedback loop to improve AI models using explanations.	User's trust and reliance on the AI system based on the provided explanations	No clear explanation method in the context of safety of human-robot interactions and/or in human-activity recognition	Experts agree with 90% of the AI decisions	To be evaluated

Federated learning-based in-vehicle Intrusion Detection System (IDS) evaluation [ODINS]

To test the performance of the FL-IDS model on the data collected during the tests, it was first necessary to make some adaptations, both to the model itself to fit the format of the simulator data (in MQTT messages), and to the data itself, since the simulator cannot simulate behaviours of attacking vehicles. Therefore, based on the original telemetry data (position and speed on the X and Y axes, which are the main variables used by our FL-IDS), synthetic data were generated simulating the behaviour of attacking vehicles of different classes, namely: Constant Position, Constant Position Offset, Random Position, Random Position Offset, Constant Speed, Constant Speed Offset, Random Speed, and Random Speed Offset, matching the attacks included in the VeReMi dataset, which is the most standardised and widely used for the task of intrusion detection in vehicles.

Going into more detail regarding the synthetic generation of the data, for the Constant Position class, which consists of the vehicle always transmitting a constant position even though its speed changes, the data generation was very straightforward. The original data were divided into 10 consecutive blocks (each containing 10% of the data), within each block, the position values on the X and Y axes of each row (hereafter referred to as pos_x and pos_y respectively), were replaced by the pos_x and pos_y values of the first record of each block, keeping the original speed data on the X and Y axes (which from now on we will refer to as spd_x and spd_y, respectively). It is worth noting that for the

generation of the Constant Speed class, which conversely consists of the attacking vehicle always transmitting a constant speed even though its position changes, the same reasoning was followed, but in reverse, keeping the spd_x and spd_y values constant within each block, while the pos_x and pos_y values were kept as in the original data.

For the simulation of data belonging to the Constant Position Offset class, in which the attacker adds a constant offset to the real position, the original data were once again divided into 10 blocks of consecutive records, each comprising 10% of the total. Then, for each block, a displacement on the X axis and another on the Y axis were randomly generated according to a uniform distribution. These displacements were respectively added to the pos_x and pos_y values of each record within the corresponding block. Once again, spd_x and spd_y remained unchanged. In turn, for the synthetic creation of the Constant Speed Offset attack, where the malicious vehicle adds a constant offset to its real speed, exactly the same reasoning was followed, but generating and adding the constant offset for each block to spd_x and spd_y, also according to a uniform distribution adapted to the range of speed values.

The Random Position attack consists of the vehicle transmitting a random position at each point in time, regardless of its real speed. Therefore, its synthetic generation from the original normal behaviour data consisted of, for each record, replacing the pos_x and pos_y values with random ones, again generated according to a uniform distribution adapted to the range of real position values of the circuit, without altering the speed. Thus, the Random Speed attack, in which the attacker sends random speeds at each moment, independently of its real position, was simulated following the same process as in Random Position, but for spd_x and spd_y.

Finally, in Random Position Offset, the malicious vehicle adds to the real position of the vehicle a random offset that changes at each point in time. Therefore, its generation was very similar to that of Constant Position Offset, but instead of generating a single offset for an entire block, in this case, a new offset was randomly generated for each record, also following a uniform distribution according to the real range of position values. The same procedure was followed for Random Speed Offset, but randomly generating the offsets for spd_x and spd_y, since here the attacker transmits a random speed at each moment.

Once the adaptation and generation of the data had been explained, it is worth noting that some slight modifications were made to the model scripts, simply so that they could work with the JSON format of MQTT messages. The same structure of the model trained on the data from the extended VeReMi dataset was reused, as described in previous deliverables, and, above all, in (Zambudio et al. 2025), consisting of an LSTM with 128 neurons and two hidden dense layers of 64 and 32 neurons respectively, with various intermediate dropout layers to prevent overfitting. Nevertheless, the size of the input layer was modified to allow the use of new predictors derived from pos_x, pos_y, spd_x and spd_y, and their respective temporal lags. The number of required rounds was also adjusted to 30, the number of local epochs per client to 10, and the batch size to 32.

For the federated training of the model on the data, it is worth noting that, since telemetry data were available for each of the participants in each circuit separately, each federated client in the training was matched with the data of one participant in the same circuit, up to a total of 40 clients. In turn, so that the server could evaluate the model's performance at the end of each round of federated training, the data of the remaining participants (not used by the clients) in the same circuit were employed.

In this way, the model was evaluated on the server's test data (independent from those of the clients), and for the first circuit the model achieved an accuracy of 88%, while for the second circuit it reached an 82% success rate, thus demonstrating good results, although there is still room for improvement,

which will continue to be worked on throughout the project. Figure 60 and Figure 61 show, respectively, the confusion matrices obtained by the model on the server's test data in circuits P1 and P2.

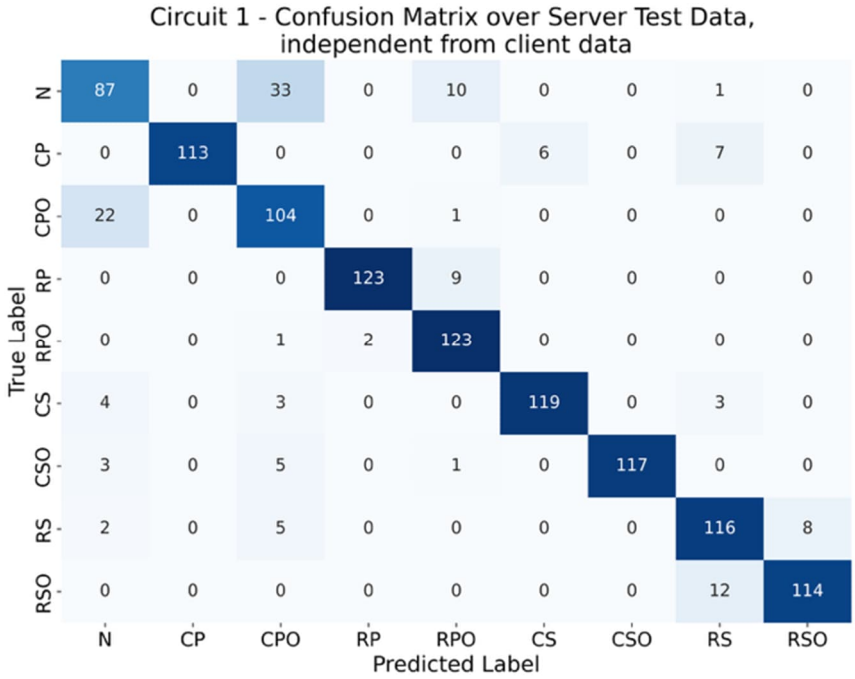


Figure 60 Confusion matrix obtained by the model on the server's test data in circuit P1 (88% accuracy)

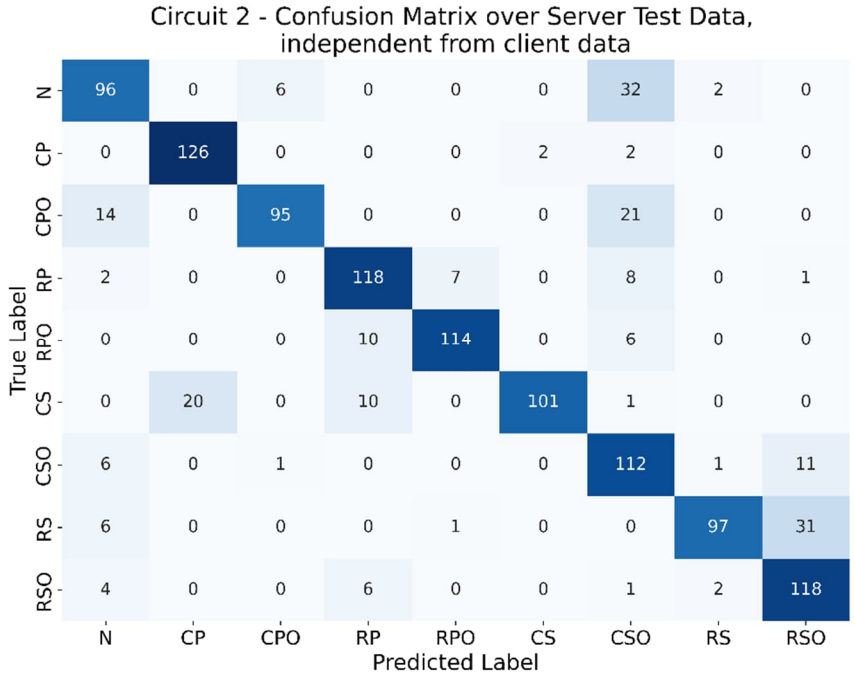


Figure 61 Confusion matrix obtained by the model on the server's test data in circuit P2 (82% accuracy)

Mental state extraction evaluation and Face-based driver monitoring [EMOJ]

Participants' facial expressions and events happened during the case study are recorded in separate csv files. Due to recorded timestamp inconsistency and system delays in recordings, data preprocessing steps are taken to align all the available data to better conduct the analysis. Facial expression analysis is recorded approximately in 0.1s interval, while the events have 0.05s interval. In facial analysis data, valence and engagement have different scales -100 to 100 and 0 to 100 respectively, while yaw and pitch are in degrees and in the allowed range of 30 and 40. The events are recorded in binary, they are expressed as 1 when events occurred and switches back to 0 when no related events occur. All these values are standardised to be expressed in -1 to 1 scale before conducting the analysis.

To better incorporate different timestamps resolutions, binning technique is used to upscale the timestamps to 1 second.

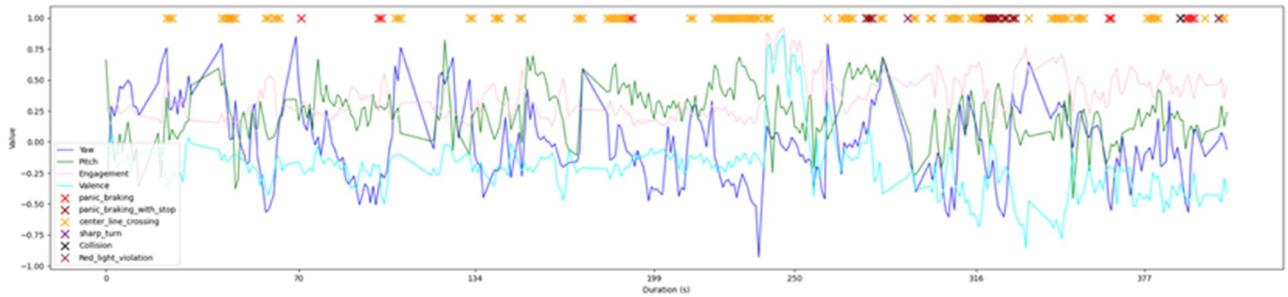


Figure 62 Drive head orientation vs. facial analysis with events

The collected data is visualised on the chart to study the relationship between driver's yaw, pitch behaviour and valence, engagement.

Yaw and Pitch: Yaw (blue line) represents horizontal rotation of the head, like looking left and right. Pitch (green line) represents vertical rotation, like looking up and down. The chart shows frequent, rapid fluctuations in both yaw and pitch throughout the drive, suggesting the driver was actively moving their head, possibly checking mirrors, looking around, or being distracted.

Panic Braking (orange X): These are relatively frequent, occurring in clusters, especially around 250s and 300s. Multiple instances of panic braking suggest the driver may have been following other vehicles too closely, or reacting to unexpected traffic conditions.

Center Line Crossing (red X): These are less frequent but still present, often occurring near instances of panic braking or sharp turns. This could be a symptom of inattention or distraction.

The chart also includes Valence and Engagement, which are psychological metrics derived from emotional data. They provide insight into the driver's emotional and mental state.

Valence (pink line): This metric indicates the emotional state, ranging from negative (-1.0) to positive (1.0). A value near 0 suggests a neutral state. The valence line generally hovers in the positive range, but with frequent, sharp drops into the negative. The most significant drop occurs around the 250-second mark, coinciding with a high concentration of panic braking. This suggests that these events caused negative emotional responses like frustration or stress.

Engagement (cyan line): This metric measures the driver's level of mental focus or concentration. A high value (approaching 1.0) indicates high engagement, while a low value (approaching -1.0) suggests a lack of focus, boredom, or distraction. The engagement line shows a pattern of erratic fluctuations. It is particularly low (below 0) during the first 100 seconds and again from around 200 to 250 seconds. This low engagement could be the underlying cause for many of the detected driving events (panic braking, center line crossing), as a less engaged driver is more likely to miss cues and react late.

The chart depicts a picture of a driver who is not consistently focused or emotionally stable while driving. The fluctuating yaw and pitch suggest physical distraction, while low engagement levels point to a lack of mental focus. The high frequency of panic braking and instances of center line crossing are direct consequences of this distracted and disengaged state. The negative valence drops during these events highlight the stressful and frustrating nature of these driving errors, creating a feedback loop where poor driving leads to negative emotions, which can further impair performance.

In this cycle, the focus was on analysing emotional states with the first version of the EMOJ's FER model trained with the public datasets CK+, Emotionet and FER+, that had achieved the following performance in terms of accuracy:

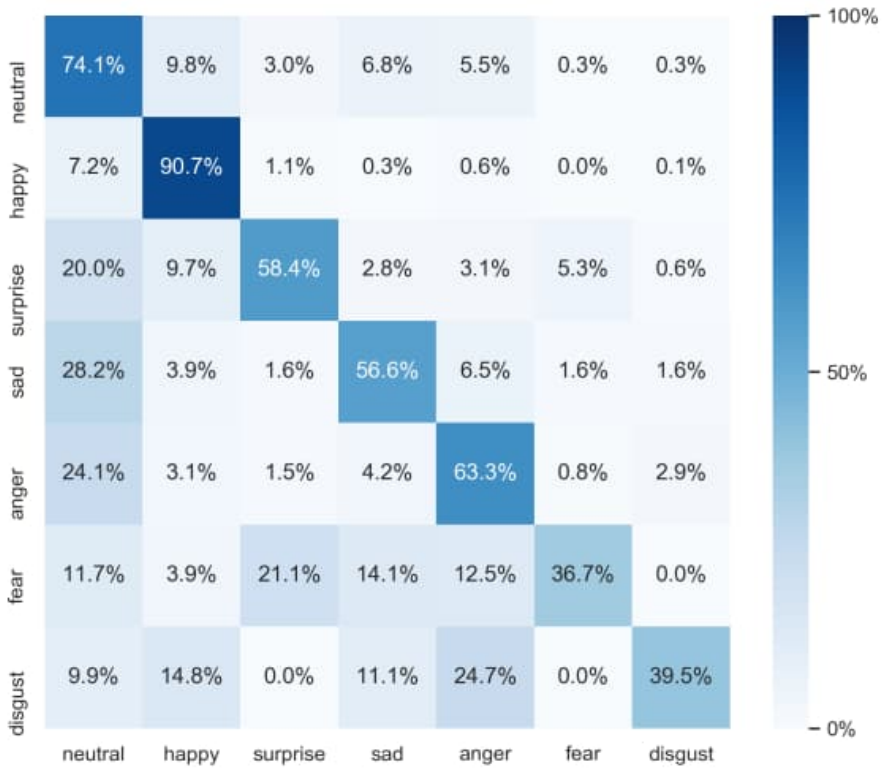


Figure 63 Confusion Matrix with accuracy values for cycle 1 FER model

These values will be compared with the accuracy values obtained from the new network based on synthetic datasets that will be developed in the second cycle.

Simulated VRUs detection [AITEK]

In the first cycle of the project, a first training was conducted using the videos generated during the data collection weeks in July in the Re:Lab premises. The videos represent the ScanER simulator during the driving of different subjects. In this case, the focus is on the detection of different classes that a car may meet on the street, such as persons, bicycles, cars, motorcycles, buses, trucks, traffic lights, and road signs, as we can see in Figure 64. The image highlights how the trained YOLO model (You Only Look Once, Unified, Real-Time Object Detection) is able to correctly identify these different classes with varying confidence scores, based on different factors, as the distance from the class of object. During this phase, we faced some class labeling challenges, since the network was not able to detect most road signs, since the default YOLO configuration only included "stop signs" and ignored other types of signs. To overcome this limitation, we extended the labeling to cover a

broader set of road signs, allowing the model to learn and generalize beyond the default conditions. In addition, in this phase we revised the labeling of some large cars, such as pickup trucks. Initially, YOLO classified them as “trucks,” but we decided to include pickups within the “car” class, reserving the “truck” label exclusively for heavy trucks. This refinement ensured a more accurate class distribution, aligned with the intended use case.

Figure 64 training set images examples showing the classes of interest

The results could be summarized in the normalized confusion matrix shown in Figure 2a. The neural network performance across most categories is good, with detection accuracy exceeding 94% for most objects. Specifically, the model achieved 0.98 accuracy for persons, 0.97 for bicycles, 0.99 for cars, and 0.94 for motorcycles and trucks. Road infrastructure elements, such as traffic lights (0.98) and road signs (0.98), were also recognized as high reliability.

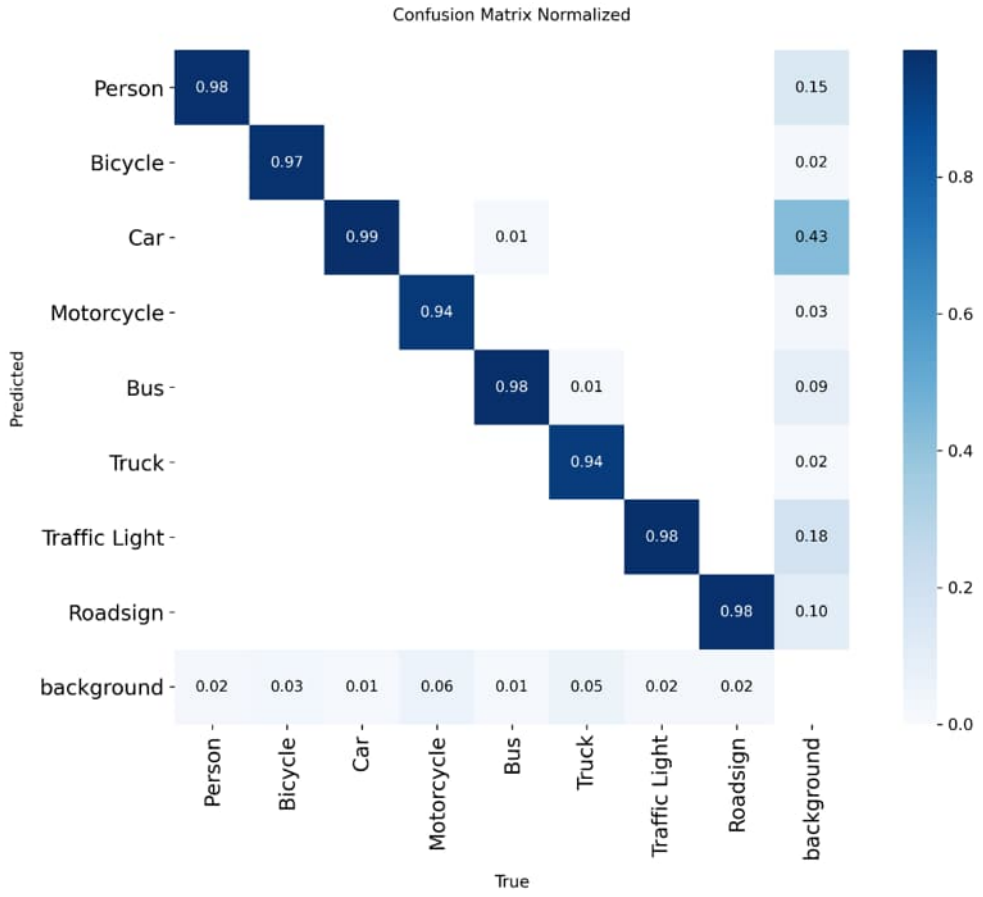


Figure 65 Confusion Matrix of the detection algorithm fo the classes of interest for the use case

Some challenges emerged in relation to background misclassification, particularly in the cases of cars (0.43) and persons (0.15) being occasionally confused with background regions. These results highlight the need for additional fine-tuning and the integration of more diverse training samples, especially under complex or cluttered environments.

Explainable AI evaluation [RULEX]

At the current stage of the project, it is not possible to perform a full evaluation of the explainability mechanisms against expert knowledge, for two main reasons. First, the activities foreseen in this task rely on the availability of data streams generated by other modules, such as models assessing driver distraction or other virtual sensors that capture critical signals for autonomous driving. These data are essential to test whether the explanations provided by the XAI (Explainable artificial intelligence) methods correctly highlight the relevant features used by domain experts in their decision-making. However, such data are not yet available, since the corresponding modules are still under development and validation in other parts of the project.

Second, the evaluation of explanations presupposes the existence of mature AI models to be explained. The design of a feedback loop, aimed at improving the models based on explanations, requires that the underlying neural networks or predictive systems (e.g., based on images, time-series, or multimodal sensor data) are already trained and deployed. At this point in time, those models are not yet finalized, and it would be premature to test the alignment between expert knowledge and algorithmic explanations.

For these reasons, the present activity can only be framed conceptually, defining the methodology and preparing the tools for future evaluation. The actual comparison between expert reasoning and model explanations will be feasible only once both the necessary data and the reference models are fully available. This phased approach ensures that the evaluation will be robust, realistic, and directly linked to the operational context of the autonomous driving use case.

Data collection for driver and vehicle via digital twins [UNIBO]

The collection of driver and vehicle related data is built around dedicated digital twins, which integrate a perception layer implemented as a docker container. This layer subscribes to selected MQTT topics to acquire real-time data streams and stores them in a MongoDB instance. Both the digital twins and the MongoDB database are deployed on an NVIDIA Jetson Orin edge node, ensuring efficient, low-latency processing at the edge. The system performance has been assessed by measuring receiving latency, defined as the difference between the time at which data are created and the time they are received by digital twins. As can be seen in Figure 66 and Figure 67, latency distributions remain consistently low in both driving with distraction and no-distraction scenarios, typically ranging between 30 and 80 milliseconds. Driving with distraction occasionally introduces higher latency peaks (up to ~0.16 s), while the no-distraction condition shows more frequent low-latency outliers. Overall, the magnitude of delays confirms that the chosen architecture—MQTT for message delivery, MongoDB for storage, and edge deployment on Jetson Orin—supports near–real-time synchronization between the physical system and its digital counterpart. The few observed peaks are likely due to transient network or processing loads rather than the distraction condition itself, reinforcing the robustness of the setup for continuous monitoring and simulation-driven testing.

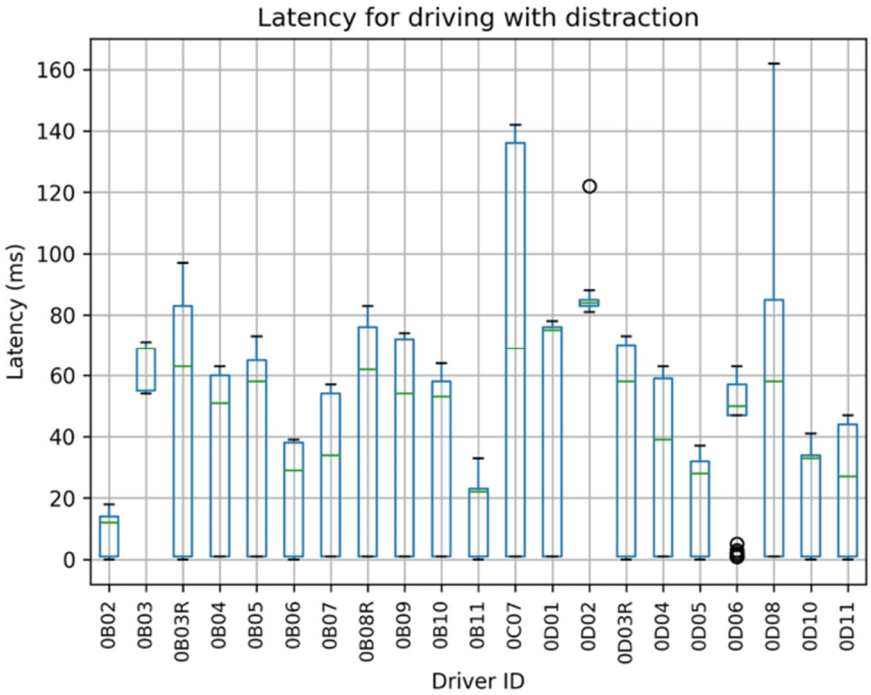


Figure 66 Latency distribution over driving with distraction scenarios

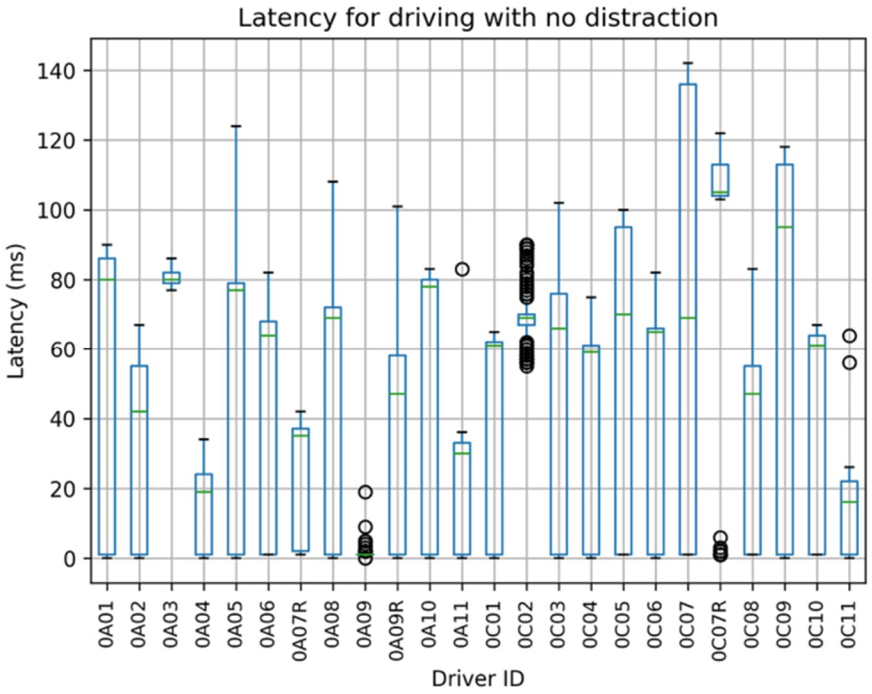


Figure 67 Latency distribution over driving with distraction scenarios

The gathered data indicates that both the software and hardware setups are sufficiently powerful to handle the generated data flow efficiently. In particular, our results demonstrate that no bottlenecks occurred, with latency distribution consistently ranging between 0.03 and 0.08 seconds. This narrow range confirms the stability of the system under load and highlights its capacity to maintain real-time performance. Consequently, the current configuration can be considered robust and reliable for the intended workload, with no immediate need for further optimization.

During simulator-based driving tests, the Coospo HW9 wearable device was employed to acquire heart rate (HR) and RR intervals, while the emotion detection and distraction detection modules (They will be developed by project partners) were integrated to complement the dataset. These data streams constitute the perception layer of both the Driver Digital Twin (DrDT), forming the basis for the driver's psychophysical state representation.

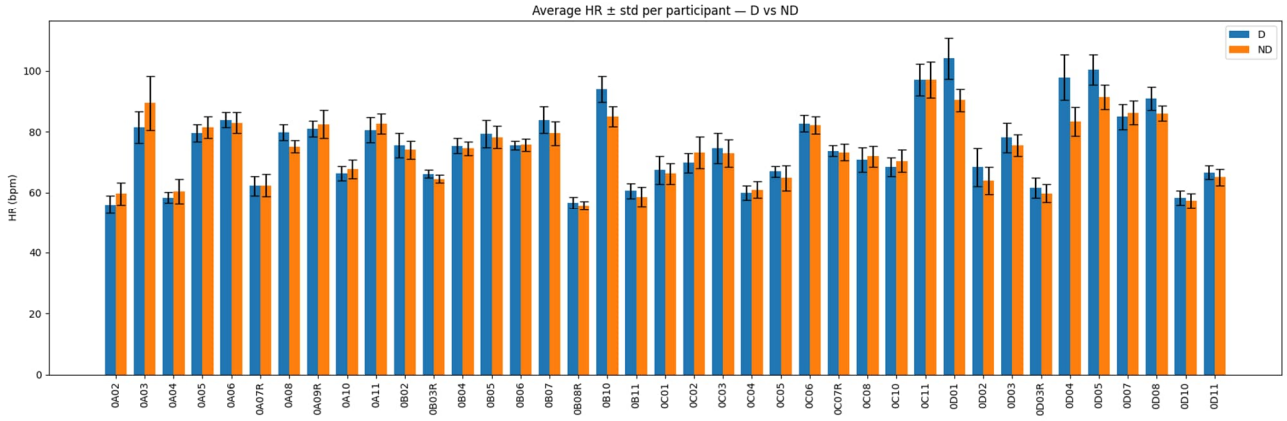


Figure 68 Average heart rate (HR) ± standard deviation per participant under distracted (D, blue) and non-distracted (ND, orange) driving conditions, as recorded with the Coospo HW9 during simulator-based tests.

Preliminary analysis has shown that RR intervals, required for a precise computation of arousal, were often unavailable due to arm movements during driving sessions, whereas HR values were consistently recorded and are reported in Figure 68. This evidence will guide the next cycle, where continual learning algorithms will be designed to ensure robustness to missing RR data and to enable accurate arousal estimation.

4.4.4 User aspects: stakeholder engagement in pilot development

For the P2-REGG pilot, four subject matter experts with consolidated expertise in human–machine interfaces and automotive innovation have been engaged. This choice was facilitated by the strong network of companies and practitioners represented within the consortium, which provided access to expertise grounded in both industrial practice and technological development.

The engagement is organised in two stages. The first, defined as formative engagement, focuses on the validation of scenarios and the collection of expert feedback relevant to the effectiveness of the final solution represented by the P2-REGG pilot.

Semi-structured interviews will be conducted, supported by carefully prepared design materials and preliminary mock-ups of the prototype under construction as well as of the results obtained during the first project cycle. This material will serve as a common reference point, enabling experts to assess whether any relevant situations or interactions have been overlooked and to situate their feedback within a concrete design framework.

Building on this introduction, the semi-structured interviews, lasting no more than 45 minutes, will address three guiding themes: the principal obstacles to adoption (including technological, market-related, social acceptance, and privacy issues), the most promising directions for future development, and the risks or design mistakes that should be avoided. The expected outcome of the formative engagement is a refined set of scenarios, accompanied by an expert-informed anticipation of adoption barriers and concerns, as well as relevant design directions.

The four experts have already confirmed their availability, and the interviews will be conducted in October 2025. Feedback will be reported in the form of design refinements, to be detailed in subsequent deliverables.

The second stage is conceived as summative engagement, which will take place once the pilot has reached a sufficient level of maturity. In this phase, experts will be re-engaged to compare their initial expectations with the actual implementation achieved by the project. Demonstration sessions with interactive HMI prototypes will be organised to enable this reflective assessment. The purpose is to gather judgments on usability, acceptance, and to evaluate how the technology addresses the risks and opportunities identified during the formative stage.

4.4.5 **User-based KPI assessment**

In the first week of July 2025, in the simulator environment set-up described above, we conducted a data collection campaign aimed at two main objectives. First, we collected behavioral data to support the training of the AI modules involved in the pilot, including those responsible for detecting the driver's state and for the decision support system designed to identify hazardous situations. Second, the tests were used to establish a baseline of user based KPI in terms of driver performance and perceived experience, which will serve as a reference point to assess the improvements introduced by the multisensory system during Cycle 2.

Namely, the considered user based KPI as presented in D5.1 were:

- Number of accidents with VRUs: total number of collisions with vulnerable road users (VRUs) during simulation, under both distracted and non-distracted conditions.
- Red Light Violations: instances in which the vehicle crossed an intersection while the traffic light was red.
- Speeding Violations: occurrences of the vehicle exceeding the legal speed limit.
- Harsh Steering: sudden and sharp changes in the steering angle, indicating abrupt or unsafe maneuvers.
- Harsh Braking with Full Stop: emergency braking events that resulted in a complete stop, possibly due to critical traffic events.
- Harsh Braking without Full Stop: sudden deceleration events where the vehicle did not come to a full stop, potentially indicating near-miss situations.
- Lane Departures: instances where the vehicle unintentionally crossed lane markings, which may reflect a lack of attention or control.
- Perceived Situational awareness: assessed via the SART questionnaire, measuring the driver's ability to perceive, understand, and anticipate traffic events.
- Perceived Workload: a composite measure of the driver's perceived workload during the driving session (NASA-TLX score), derived from the weighted combination of six subscales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration). It reflects the overall cognitive and physical strain experienced by the user.
- Perceived User experience: evaluated using the short version of the User Experience Questionnaire (UEQ-s), focusing on usability, engagement, and emotional response.

The data collection protocol was introduced in the previous deliverable (D5.1) and is further detailed in a dedicated publication to appear in the proceedings of the HCII 2025 conference (Presta et al., 2025, in Section 7). The protocol is briefly recalled here to facilitate the readability of the contribution. The simulation includes five categories of risk scenarios involving vulnerable road users (VRUs), each reflecting a common urban situation where driver awareness is likely to be challenged. For each category, the simulation presents four variations: two critical events (where the VRU crosses the driver's path), one non-critical event (with a visible VRU not interfering), and one neutral event (with no VRU). This results in 20 events per driving session, including 10 truly critical ones, two for each scenario type. To preserve ecological validity and avoid response bias, variations were designed to create ambiguity and prevent participants from forming expectations. The simulation includes two different route sequences (P1 and P2) and two driving conditions: non-distracted (ND) and distracted (D), with the latter introducing a secondary visual research task (6 tasks per minute). To control for order and sequence effects, participants were assigned to one of four experimental groups. Each group completed two driving sessions, one in the ND condition and one in the D condition, combined with different route sequences. Specifically, Group A followed the sequence ND–P1 in the first session and D–P2 in the second; Group B experienced D–P1 first, followed by ND–P2; Group C completed ND–P2 first and D–P1 second; and Group D began with D–P2 and concluded with ND–P1. This balanced rotation ensured that all combinations of route and condition order were equally represented across participants, thus minimizing potential learning, fatigue, or anticipation effects. At the end of each session, participants completed an exit questionnaire to measure the perceived experience and at the end of both the driving sessions they answered a short interview. An initial questionnaire collected demographic information and driving experience.

The data collection took place in the driving simulator at RE:Lab with 48 volunteer participants. All participants were informed about the study both verbally and in writing by the experimenters and gave their consent by signing an informed consent form prior to participation. Due to technical issues, the final sample consisted of 41 participants (15 female, 26 male) with a mean age of 33.2 years (SD = 8.4). Regarding driving experience, 3 participants reported less than 3 years of driving, 16 had between 3 and 10 years, and 22 had more than 10 years. All participants were informed about the study procedures and provided written informed consent prior to participation. For self-assessment measures, data from one participant were excluded due to incomplete questionnaire responses, resulting in a sample of 40. Table 19 reports the per-participant average results for the user-based KPIs measured during the baseline assessment.

*Table 19 Average per-participant results for user-based KPIs measured during the baseline assessment. Values are reported as overall means (Mean tot) and separated between **distraction (D)** and **non-distraction (ND)** conditions. The KPIs include safety indicators (accidents with VRUs, red light violations, speeding violations, harsh steering, harsh braking with and without full stop, lane departures) and subjective metrics (perceived situational awareness, perceived workload, perceived user experience).*

User-based KPI	Mean (tot)	Mean (D)	Mean (ND)
Number of accidents with VRUs	2.20	0.73	1.46
Red Light Violations	0.29	0.2	0.1
Speeding Violations	10.68	5.1	5.59
Harsh Steering	7.63	3.73	3.9
Harsh Braking with Full Stop	1.29	0.71	0.59
Harsh Braking without Full Stop	1.07	0.41	0.66
Lane Departures	49.39	26.24	23.15
Perceived Situational awareness	4.33	3.82	4.85
Perceived Workload	3.41	3.07	3.75
Perceived User Experience	0.256	0.218	0.295

Among the user-based safety metrics, only two indicators show statistically significant differences between the distracted (D) and non-distracted (ND) driving conditions. This finding is based on a paired-sample t-test conducted across all participants. First, the number of collisions was statistically lower in the distracted condition. While this result may seem counterintuitive, it could be attributed to compensatory behaviours, such as increased caution or reduced speed, when participants were aware of being cognitively loaded by the secondary task. Second, the number of lane departures was higher in the distracted condition (M = 26.24) than in the non-distracted one (M = 23.15). This result suggests that visual distraction may impair lateral control and increase difficulty in maintaining lane position, even when other behaviours remain unaffected.

On the other hand, looking at the perceived experience metrics, the overall perceived workload, measured through the aggregated NASA-TLX score, was significantly higher in the distracted condition compared to the non-distracted condition. The difference was statistically significant. This result confirms that the introduction of the secondary task during driving led to a measurable increase in subjective workload, affecting participants' overall perception of task demand and mental effort. The overall SART index was significantly lower in the distracted condition compared to the non-distracted condition. This result indicates a measurable decline in participants' perceived situational awareness under distraction, suggesting that the secondary task reduced their ability to monitor and integrate information from the driving environment.

The UEQ-S analysis showed that the overall user experience was perceived as neutral in both conditions (M=0.256), with a slight decrease under distraction (M(D) = 0.218, M(ND) = 0.295). This pattern is consistent with the higher workload and lower situational awareness reported in the distracted condition.

The data currently being used to support the development of the technical modules effectively capture these main behavioral and perceptual patterns.

Taken together, these results suggest a nuanced interaction between cognitive load, behavioral adaptation, and situational awareness. The statistically lower number of collisions observed in the distracted condition may initially appear paradoxical, especially when considering the increased workload and reduced perceived situation awareness reported by participants. However, this discrepancy can be interpreted in light of compensatory mechanisms often observed in human performance under dual-task conditions. When faced with an additional cognitive challenge, such as the visual secondary task introduced in the distracted condition, participants may have consciously adopted a more conservative or risk-averse driving style, prioritizing safety and devoting more deliberate attention to the driving task despite the increased mental effort required. This is consistent with the elevated NASA-TLX scores, which reflect the subjective strain of maintaining performance under load, and with the reduced SART index, indicating diminished perceived control and environmental comprehension. In contrast, the non-distracted condition, although not inherently less demanding (as it involved complex urban driving with active navigation), may have induced a more fluid but less guarded driving mode, possibly leading to a higher rate of collisions. Thus, the data suggest that while distraction impairs certain aspects of vehicle control (as evidenced by increased lane departures), it may also trigger a compensatory increase in task engagement that temporarily mitigates some types of risk, particularly those requiring forward attention and proactive hazard avoidance.

When confronted with a secondary cognitive task, such as the visual distraction introduced in the D condition, participants may have consciously adopted a more cautious and deliberate driving style. This compensatory engagement, aimed at maintaining safety despite increased mental demand, is supported not only by the behavioral data but also by participants' self-reports in post-session interviews. Several participants explicitly acknowledged that they felt more focused or alert during the distracted condition, precisely because they were aware of the additional challenge.

Moreover, this effect may have been reinforced by characteristics of the experimental setting. Participants were instructed to maintain a safe driving style in both conditions and were aware that their performance was being observed and recorded. As a result, some may have intuitively understood that avoiding collisions was a central focus of the study, leading to increased vigilance, particularly when they felt cognitively compromised. Another contributing factor may lie in the nature of the distraction task itself, which some participants found unexpectedly engaging or stimulating, possibly enhancing their overall arousal and commitment to the driving task.

Thus, while distraction clearly impaired certain aspects of vehicle control (as shown by the increase in lane departures), it may also have triggered a compensatory increase in task engagement and vigilance, temporarily mitigating risks that rely on forward attention and proactive hazard detection. These findings also raise the opportunity to reflect on the characteristics of the distraction task itself, which being experimenter-controlled, could be adapted in future iterations to test different levels or modalities of cognitive load, if needed.

Taken together, the significant findings, including the increase in lane departures, the higher perceived workload, and the drop in situational awareness, paint a clear picture of a cognitively overloaded and potentially unsafe driving condition. Although the number of collisions was lower in

the distracted condition, this outcome must be interpreted with caution: it likely reflects a temporary and mentally costly compensatory effort, rather than a genuine improvement in safety. In fact, the overall scenario remains critical, with multiple indicators suggesting that the driver's ability to monitor and control the environment is compromised. This is particularly relevant given that even in the non-distracted condition, several participants still failed to detect or avoid vulnerable road users, an outcome that is far from acceptable in real-world driving contexts.

4.4.6 User aspects: Gender/age issues and ethical concerns in P2-REGG

For the ethical reflection on pilot P2-REGG, six participants (two women and four men), all directly involved in the management or development of the pilot, contributed on a voluntary basis.

Their roles included algorithm developers, HMI designers, human factors researchers, R&D project managers, and a strategic lead with executive responsibilities. The group brought together expertise in digital twins, computer science, computer vision and machine learning, user-centred design, and user experience.

Each contributor submitted their reflections individually via a shared form, focusing on aspects relevant to their specific area of expertise or involvement in the project. The partner in charge of compiling the Ethics Exercise form (UNISOB) then prepared a summary that brought together the different perspectives, highlighting recurring themes and complementarities among the inputs. This synthesis was subsequently shared with the P2-REGG group for validation and final adjustments, ensuring that the resulting document reflects a shared understanding among all contributors.

4.4.7 P2-REGG transition into cycle 2: takeaways and feedback

- Highlights of cycle 1
  - Successful deployment of the perception layer for both DrDT and VDT on the NVIDIA Jetson Orin edge node, with efficient integration of MQTT for data streaming and MongoDB for storage.
  - Latency evaluation confirmed stable near–real-time performance (30–80 ms on average), demonstrating the robustness of the architecture under both distracted and non-distracted driving conditions.
- Lowlights of cycle 1 evaluation
  - Limited focus so far on higher layers of the DrDT and VDT (beyond perception), as efforts were primarily dedicated to data collection, synchronization, and initial validation of the architecture.
- Intended changes for cycle 2 specification and evaluation of P2-REGG
  - Extension of the DrDT and VDT beyond the perception layer towards the full digital twin pipeline, including processing, decision-making, and interaction layers.

4.5 P2-TMP evaluation cycle 1

The pilot P2-TMP focuses on driver monitoring in urban traffic environments: collecting driver monitoring information from inside the vehicle and combining it with external sensor data collected from outside of the vehicle. This pilot is made up of two demonstrators: demonstration 2.2.2 is focused on monitoring the attention and emotions of the vehicle driver and passengers, while demonstration 2.2.3 aims to match driver's attention to the traffic situation around the vehicle, to estimate how aware of their surroundings the driver is during driving.

The first cycle of the project was planned to consist of collecting internal and external sensor data for all pilot partners and potentially creating an open-source driver monitoring data set for public use. This collected data was to be used for offline analysis and algorithm development in preparation for the second cycle. In the second cycle, the objective is to further develop these algorithms and bring them online to use in real traffic situations to add to the driver's situational awareness.

The pilot uses VTT's research vehicle Heluna (Volkswagen Multivan) for online algorithm testing and data collection. The data collection and online tests are conducted in the city of Tampere: the test area covering both busy urban streets and less hectic passing street sections. The vehicle is equipped with internal and external sensors for in-vehicle and off-vehicle monitoring, as well as edge computing devices for running the software logic modules and AI algorithms.

4.5.1 Final pilot set-up

The pilot set-up has not changed since the initial planning described in deliverables D5.1 and D5.2 as the test drives are yet to be completed. The following description thus follows the one given in said deliverables.

Data collection rides will take place in the Tampere city region, on public roads, always following the same predefined route. The planned route is shown in Figure 69. It begins on a primary urban road, passing the city center. This part includes a long tunnel drive, so more challenging lighting conditions are included in the data sets. After this, the route moves onto smaller roads inside the city center, where traffic includes plenty of VRUs. The route includes different speed limits.

The route is driven by different test subjects, chosen from withing VTTs employees. Each driver is briefed on what data is collected of them and for what purpose. Each driver agrees to the terms of data collection and sharing, before any tests are run. The test subjects are briefed on the driven route, and additionally a navigator will instruct the driver during the ride. A researcher is also present in the vehicle and will provide more instructions when necessary. The driver is given a health monitoring belt during the drive, that collects data of their heart and respiration rate. They are instructed not to speak (unless asking for instructions) during the drive, so as not to interfere with these measurements. After the route is finished, they will switch places with another participant and take the role of passenger. No tasks to measure distraction are required by the driver, so they can fully immerse themselves in the task of driving safely. There are also no tasks or requirements for the passengers, and they are allowed to talk to each other, as long as they don't disturb the driver excessively.

The pilot's sensor set-up includes the following:

- 2 RGB cameras inside the vehicle: one pointed straight toward the driver's face, and another pointed from the side, catching images or both the driver and the passenger on the front seat. No additional lighting, as the cameras determine their exposure times automatically.
- 1 RGB camera located on top of the vehicle for detecting objects from the traffic ahead to the vehicle.
- LiDAR sensor outside the vehicle for generating 3D point cloud imagery of the surroundings, used for detection and tracking of other road users.
- Wearable sensor device (Zephyr), which provides a reference signal for physiological measurements.

Other data collected from the vehicle, not directly relevant for driver monitoring, but useful for further analysis:

- IMU (inertial measurement unit), vehicle pose
- GNSS (global navigation satellite system), vehicle's exact location
- Odometry, vehicle speed

The collected data will include cameras' intrinsic calibrations and sensors' extrinsic calibration values in relation to the vehicle's coordinate system. All data will be synchronized and include timestamps. The data is stored on VTT's network drives with access restricted only to pilot partners that need the data for their activities.

The number of test subjects is expected to be 2-8 people, where each person drives the route once, and sits as a passenger 1-3 times.

The data collection is yet to be conducted, as finalising the data security and privacy related paperwork has taken significantly more effort and time than originally anticipated.

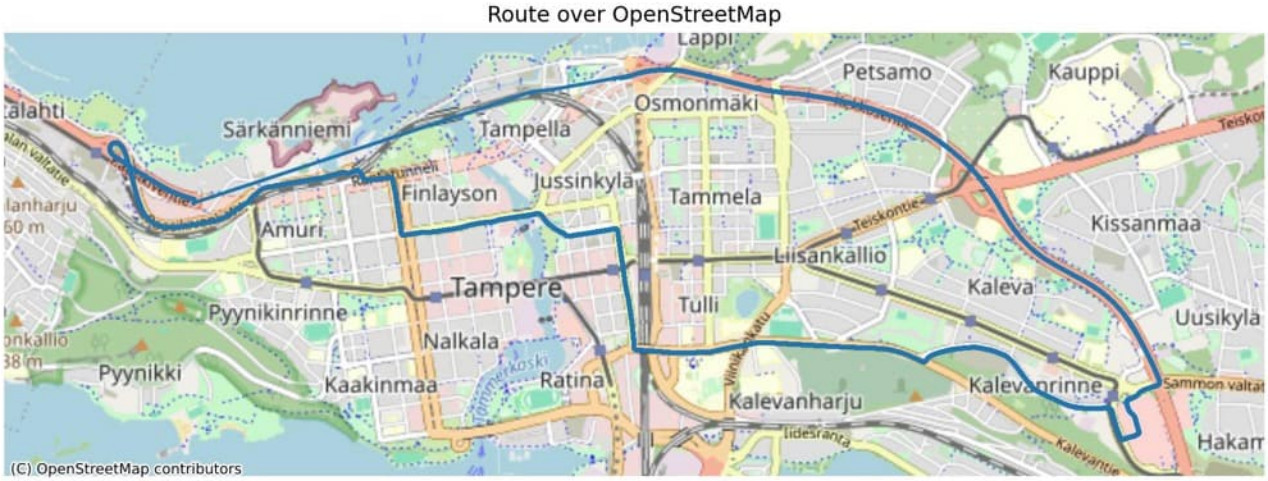


Figure 69 Test route in Tampere city

4.5.2 Pilot evaluation execution and protocol details

4.5.2.1 Pilot architecture description

The pilot's technical architecture (Figure 70) consists of three layers: the perception layer, the data processing layer, data fusion, and the final feedback to the driver and the vehicle's automated functions. The perception layer receives sensor data flowing from inside and outside the vehicle and the sensor drivers. This layer also synchronizes the data so it's all in the same time frame. The data processing layer extracts interesting information from the data with dedicated algorithms and signal processing methods, such as the vehicle driver's emotion and posture, and surrounding VRU's features. This information is then combined via data fusion to create a more holistic understanding of the traffic situation both inside and outside the vehicle, as well as matching information about both together. Cycle 1 focused on creating working perception and data processing layers, so all partners can move on to algorithm development with usable data sets.

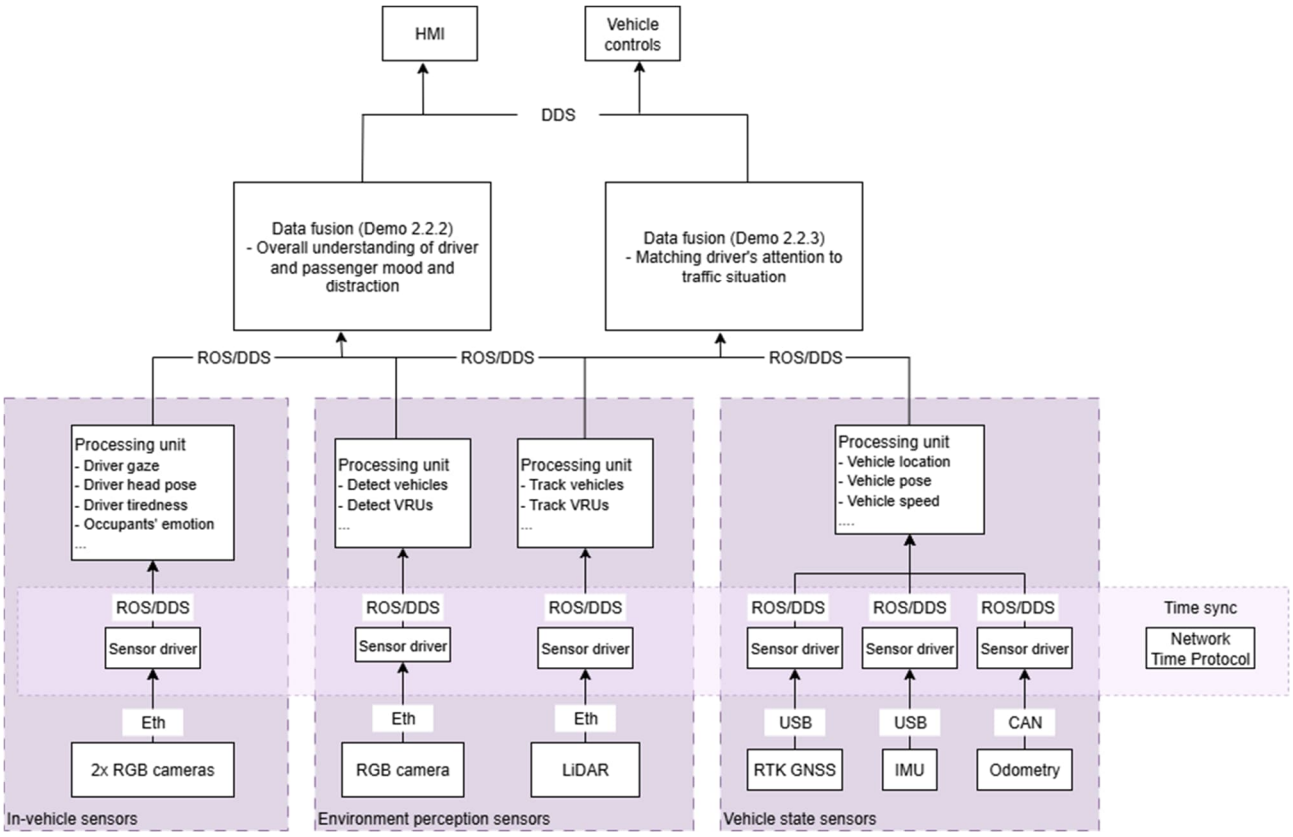


Figure 70 Updated high-level sensor and system architecture of P2-TMP pilot

4.5.2.2 Point of gaze estimation

Alongside data collection activities, VTT has developed a point of gaze estimation algorithm. Point of gaze (POG) refers to estimating what the driver is looking at (e.g. VRUs, other vehicles) around the vehicle. The system architecture of this setup is described in Figure 71.

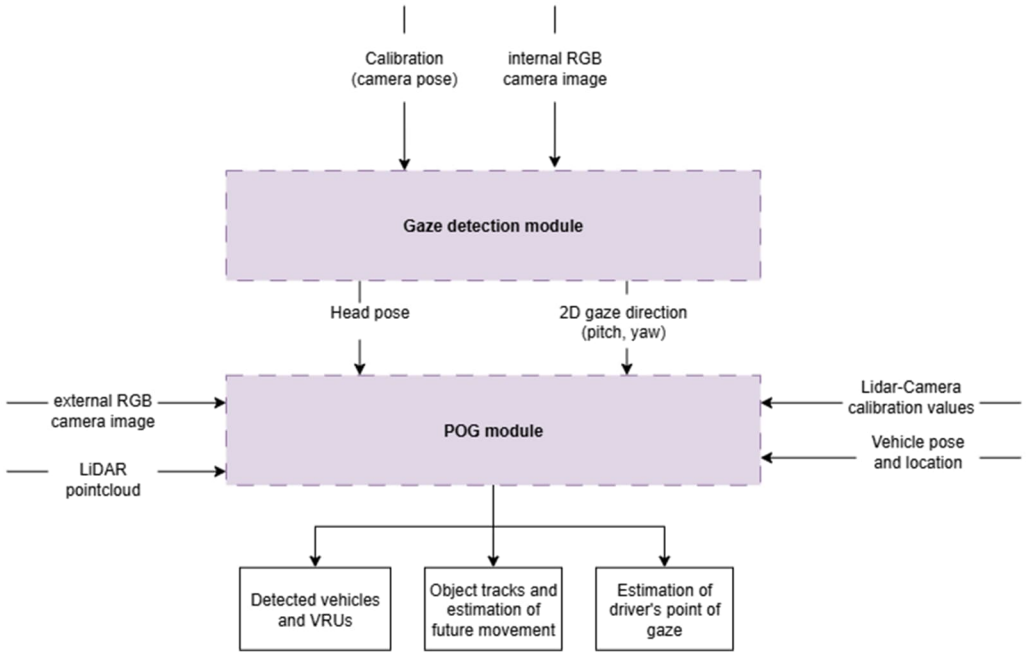


Figure 71 High-level architecture of the point of gaze estimation software

The system is made up of two modules: the gaze detection module and the POG module. The gaze detection module uses the camera's measured extrinsic pose information and camera imagery in AI algorithms to estimate the driver's head pose in relation to the camera, recognizes the driver's face, and estimates where their gaze is directed in 2D.

This information is then fed to the POG module along with external sensor imagery (LiDAR and RGB camera), measured LiDAR-camera calibration matrices, and the vehicle's status information (pose, location, odometry). VRUs and other road users are detected from the external camera view, and matched to the LiDAR point cloud, so their exact location and size can be measured. The driver's gaze is projected into a 3D gaze vector, and matched to the LiDAR view, so that the object that is pierced by the gaze vector is estimated to be what the driver is looking at. Further on, the VRUs are tracked and their future movements predicted, so possible collisions can be predicted, too.

Estimating the driver's point of gaze adds traffic safety when combined with external sensor's object detection capabilities. It can be used to estimate whether the driver has seen the VRUs approaching the road or pedestrian crossings. This way, the driver can be warned or the vehicle automatically slowed down when a collision is in danger of happening.

The current HMI/GUI of the software is pictured in Figure 72.

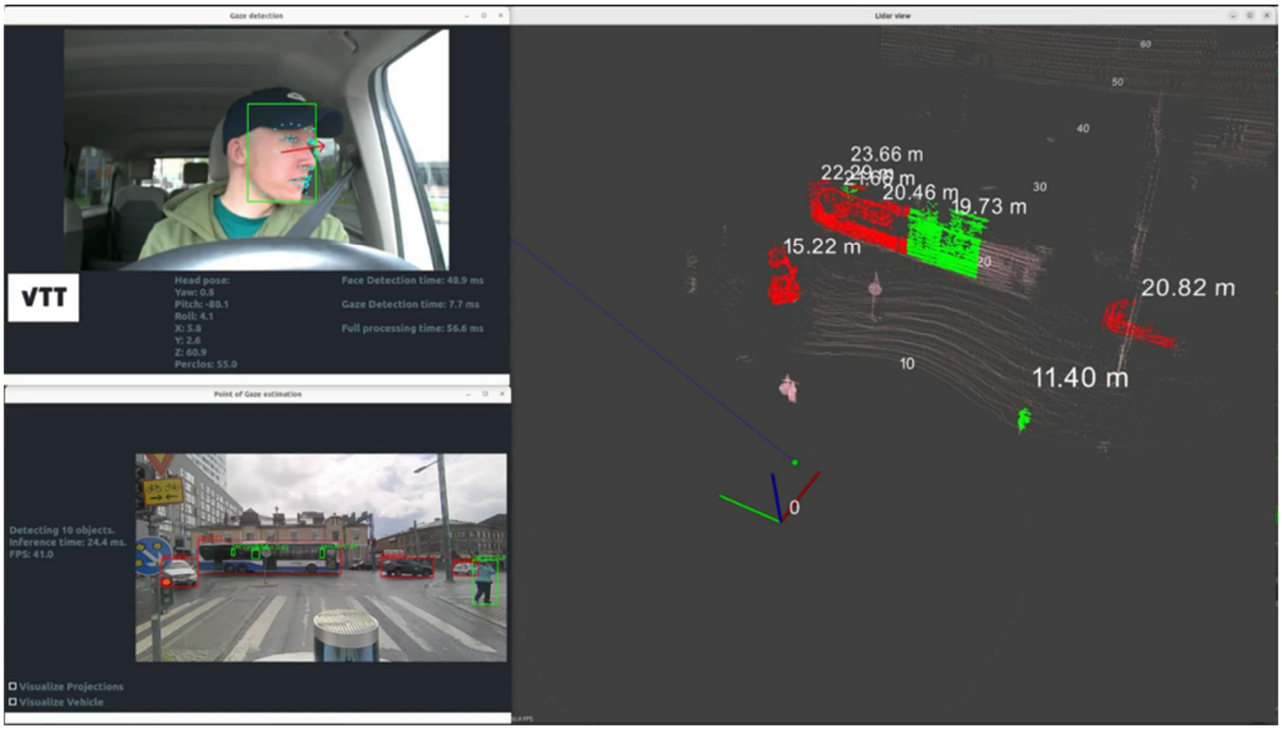


Figure 72 Point of gaze software: driver monitoring, camera detection, and lidar matching

In this, the internal camera image is pictured in the upper left corner. The faze detection is visualized as a green box, the facial landmarks as blue dots, and the 2D gaze estimation as a red arrow. The lower left image shows the external camera image with object detections visualized as boxes around them. The right side shows the LiDAR point cloud, with the matched object detections coloured in green and red. The projected 3D gaze vector is visualized as a blue line, starting from the estimated head location in relation to the vehicle's rear axle.

4.5.3 Pilot technical KPI measurements

Table 20 Technical KPIs assessed in pilot P2-TMP.

Nr	State of the art	Innovation beyond SotA	KPI	Start state	Target
4.2.1	**4.2 Mental state extraction** – Relies on the reliability and quality of the psychophysiological and behavioural signals recorded during a period of interest. In real life context, mental state detection relies on heart rate dynamics and in some cases, skin conductivity variation. Also, micromovements, gestures, group dynamics (positions and distances of persons in a group) and facial expression analysis can be used to complement the measurements.	We aim to improve understanding of the relationship between the various signals and subjective mental state.	Mental state recognition in a real-life setting	Depending on scenario. Arousal detection depends on physiological measurement (see 4.1). Valence largely on lighting conditions (face)	Accuracy improvement of 20%
4.7	**4.7 Face-based driver monitoring** – There exist systems for face-based monitoring[6] but there are limitations, for example, it is difficult to analyze face material from challenging camera angles or in a low-light scenario.	We intend to improve the ability of AI to analyze faces visually in various circumstances, which may involve challenges such as low light, difficult camera angles, or previously non-detectable facial expressions.	Maximum yaw (horizontal displacement) of the face, still allowing the system to extract facial features so that the output of the extraction is adequate for further AI-based processing steps	Up to 30 degrees of yaw of the face allowed	Up to 40 degrees of yaw of the face allowed

So far, the technical KPIs have not been evaluated in the pilot as the actual test drives and data collection have not been done yet. Therefore, no changes to the assessment methodology described in D5.1 have been made.

4.5.4 User aspects: stakeholder engagement in pilot development

Stakeholders actively involved in the pilot are private drivers, practically those participating in the pilot test. In cycle 1 it was not planned to ask their feedback systematically about the use and obtrusiveness of the monitoring system and sensors. If the participants give any remarks about the system during the test drives, those are however noted and considered in the development of the system during the 2nd cycle. Additionally, more systematic methods to collect the user feedback about the system are planned via a questionnaire during the 2nd pilot phase.

4.5.5 User-based KPI assessment

User-based KPIs related to P2-TMP pilot includes obtrusiveness and robustness of the system. Aim is to have the system to be as unobtrusive as possible which can be achieved with sensor choices (non-wearable and not emitting visible light nor light possibly compromising eye health) and design of warnings and notifications. The warnings and notifications provided by the system are to be adjusted so that they do not cause irritation and ignorance with the user. The system should also function reliably in different environmental conditions (mainly varying lighting) as well as with all kinds of users regardless of their appearance. Number of accidents and risky manoeuvres are recognised as measures for driver distraction (KPI 1.4.1). Those are to be registered by the conducting researcher onboard of the vehicle during the test drive. In cycle 2, plan is to compare number of such incidents during the test drives with and without the assisting system that provides the driver warnings and notifications about the traffic situation and other road users they may have missed.

4.5.6 User aspects: Gender/age issues and ethical concerns in P2-TMP

The ethics focus group discussion for P2-TMP was held in 26th of May 2025 with participants from pilot partners EMOJ, UNITO, VALOSSA and VTT as well as the Ethics team from VTT. One major issue with the monitoring methods used in the pilot is the possible privacy deprivation of both the occupants in the vehicle and other road users outside the vehicle. Additionally, bias of the AI methods used in the pilot and its effect on the inclusiveness will be addressed in cycle 2.

4.5.7 P2-TMP transition into cycle 2: takeaways and feedback

- Highlights of cycle 1
  - Installment of sensors and computing units on-board of the test vehicle with time-synchronized data collection from all data sources.
  - Real-time assessment of driver's POG matched the external traffic scenario.
  - Careful consideration of the data security and privacy aspects related to the personal information collected in the pilot.
- Lowlights of cycle 1 evaluation
  - Constructing the Data Protection Impact Assessment (DPIA) caused a significant delay in executing the pilot test drives and data collection which affected pilot partners' possibility to evaluate their systems in time for this deliverable and cycle 2.
- Intended changes for cycle 2 specification and evaluation of P2-TMP
  - Original plans for cycle 2 still apply: installing the pilot partners' monitoring systems on the test vehicle and running them in real-time. Additionally, number and diversity of test subjects are to be increased.

5 UC 3 demonstrator evaluation cycle summary

Use Case 3 focuses on developing an intelligent, certifiable safety system that enables safe and efficient human-robot collaboration (HRC) in industrial environments. The core innovation is a modular system capable of detecting humans, recognizing their posture and body parts (especially upper limbs), and dynamically adjusting the robot's behaviour (e.g., speed or emergency stop) based on context and proximity. This system integrates sensor fusion, machine learning (ML), and explainable AI (XAI) to support real-time risk assessment and ensure compliance with safety standards (e.g., ISO 10218, ISO/TS 15066).

UC3 includes three pilots, each addressing different stages and environments of human-robot interaction. Demonstrators 3.1 (linked to P3-BEST) and 3.2 (linked to P3-GRA) focus on developing robust perception algorithms for detecting human presence and posture in shared workspaces. As both these pilots address human safety in working with robots, the focus in this first cycle of the project is on P3-GRA which is the virtual representation of the same setup of P3-BEST. Demonstrator 3.3 (linked to P3-SON) integrates the entire safety system (perception, AI, control logic, and HMI) in a real industrial setting where mobile robots work safely alongside humans.

5.1 P3-BEST evaluation cycle 1

5.1.1 Final pilot set-up

The demonstrator is setup around a palletizing robot collaborating with human operators. The setup, shown in Figure 73, includes a UR30 robot that has the ability to lift 30 kg of payload (including the gripper). To use the robot in its collaborative mode, the operating speed needs to be reduced to ensure that any forces exerted on the operator do not cause injuries.

In current practice, operator detection is typically done through laser screens and rotating laser sensors. These sensors are incapable of understanding the scene and will therefore bring the robots to a stop at any movement detected in a range of about 3 meters from the robot. Since this could limit the uptime of the robot too much, a more intelligent solution is required.

In this pilot, we will integrate multiple sensors (camera's, IMU's, robot sensor), sensor fusion and decision making, to enable the robot to move on full speed while the situation is determined to be safe and to automatically adjust its behaviour when an operator is present in the working area of the robot.

Figure 73 Image of the scenario, before the integration of DistriMuSe sensors.

Figure 74 shows the functional diagram of the demonstrator. Sensor output is first pre-processed to obtain concrete information on the operator position and state, as well as the system state and its environment. This data is then collected in the robot world model, synchronized in time and space. Next, AI algorithms determine the overall state and – more importantly – the appropriate action that needs to be taken to maintain a safe situation. These actions are then executed on the robot. The HMI feeds back on these decisions to the operator.

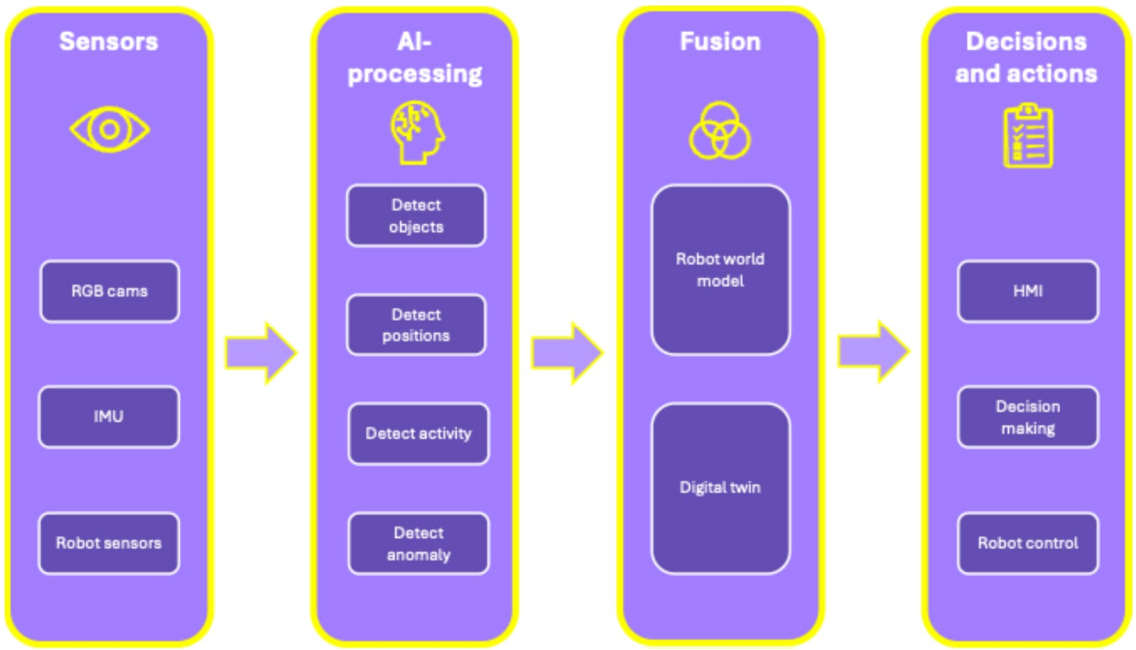


Figure 74 Functional diagram of the pilot

In this pilot, we will be working with different sensor modalities: 3 RGB cams positioned around the robot, IMU's attached to the operator body, and embedded robot sensors measuring position, speed, acceleration and force.

The positioning of the RGB sensors is indicated in Figure 75. We have 3 cameras on three different sides of the robot overseeing the robot and its direct environment. We have selected an Alhua camera of type DH-SD22404T-GN. This is a 4MP 4x PTZ Network Camera, with 1/3" 4 Megapixel CMOS and 4x optical zoom running at a maximum frame rate of 25/30fps@4M/3M.

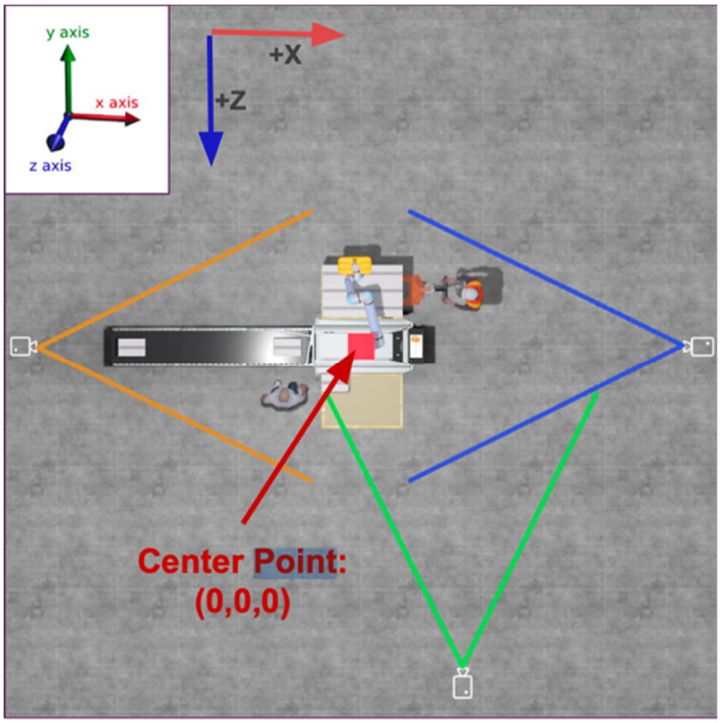


Figure 75 Positioning of the RGB cameras

5.1.2 Pilot technologies

This demonstrator integrates and tests the following technologies of the project:

5.1.2.1 Human posture detection and detection of abnormal behaviour by Pumacy

Pumacy has successfully completed the initial phase of its posture and abnormal behaviour detection system, which utilizes a network of wearable IMU sensors to capture human movement in collaborative work environments. The system was validated through extensive BLE testing, during which six sensors streamed data at 100 ms intervals over multiple hours. BLE connectivity remained highly stable, with over 99.8% of packets received within the expected interval range. This allowed for more than 8 hours of continuous, valid data collection under real-world conditions.

The initial dataset includes core human activities such as walking, standing, sitting, bending, lying down, and transitional movements. This raw sensor data has been carefully analysed and pre-processed. Key features have been extracted from the accelerometer and gyroscope signals, and the data has been structured into synchronized, labelled time-series segments suitable for machine learning.

At this stage, the focus has been on validating the end-to-end AI pipeline. A smaller subset of the collected data is being used to train and evaluate a preliminary LSTM-based deep learning model.

The goal is to ensure the correctness and effectiveness of the data processing, feature extraction, and modeling components before scaling further.

This initial training phase also enables the team to fine-tune the architecture and evaluate model behavior in offline conditions. Once this base model demonstrates satisfactory performance, a second, more comprehensive data collection campaign will be conducted. The expanded dataset will include a wider range of activities, body types, and conditions to improve model robustness and accuracy.

The final model will be optimized for deployment, enabling real-time inference of posture and abnormal behavior at the edge. This technology will serve as a core module within the demonstrator, providing critical safety and awareness capabilities in human-robot collaboration environments.

5.1.2.2 Body joint tracing, ergonomic evaluation and activity recognition by EMOJ

Emoj's Deep Learning-based system replaces the conventional "analytical" pipeline, where joint angles extracted from RGB-based Motion Capture software are plugged into formulas (e.g., from RULA, REBA, OCRA approaches), with an inductive, learning-based approach designed to be robust to perspective distortion. Analytical pipelines degrade when subjects rotate relative to the camera, because they estimate apparent (projected) rather than true ergonomic angles unless calibration is used. Humans, by contrast, judge postural risk holistically. We emulate this holistic judgment with an inductive RGB-Motion Analysis System (iRGB-MAS), which predicts ergonomic scores directly from pose keypoints without computing angles explicitly. The iRGB-MAS still relies on standard 2D pose estimators (OpenPose and MediaPipe) to localize body keypoints but then feeds all keypoints jointly into a supervised model that outputs the ergonomic index. To train it, we built a fully synthetic dataset using a controllable digital mannequin posed to cover the complete set of assessable RULA postures (excluding wrists due to RGB tracking limits). RULA categories included five for shoulder flexion/extension, two for shoulder abduction, three for elbow flexion, three for elbow flexion/extension, and four for neck and trunk flexion/extension. Enumerating these yielded 14,400 unique 3D posture configurations per mannequin. Each posture was rendered from 24 viewpoints to model arbitrary camera placement, and the resulting keypoint coordinates were duplicated with subtle perturbations to simulate detector uncertainty and partial occlusions. The final training corpus contained 1,036,800 labeled records (14,400 postures × 24 views × 3 augmentations), each annotated with its RULA-ground-truth score. We trained a Multi-Layer Perceptron (hidden_layer_sizes=100, max_iter=3000, activation=ReLU, solver=Adam) to map keypoints to ergonomic risk levels.

Alongside ergonomic scoring, the system includes a lightweight activity-recognition module grounded in anthropometric proportions. Using the relative vertical positions of keypoints for shoulders, hips, and knees, it distinguishes:

- standing/walking (hips ≈ 63% and knees ≈ 32% of shoulder height, with ±10% empirical tolerance);
- generic activity (upright but violating standing constraints due to motion or bending);
- seated (knee-to-hip distance dropping below 50% of shoulder-to-hip distance, derived from a 72% anthropometric ratio);
- picking up (hands descending below a threshold set slightly below the knees, ~110% of knee level);

When required keypoints are missing, the sample is assigned to an occluded class, with minimal visibility conditions enforced (sternum, ≥1 hip, ≥1 knee).

5.1.2.3 Event detection and localization by AITEK

The fallen person detection task has been framed as an image classification problem, which is one of the most fundamental challenges in computer vision. Unlike bounding box detection, image classification is computationally lighter and requires fewer resources, making it ideal for scenarios where real-time performance is essential. This is particularly relevant in this use case, where the ultimate objective is to guarantee the safety of human operators working in close collaboration with robots. In such contexts, rapid and reliable detection of dangerous situations, such as a person fainting close to the robot, is crucial for triggering timely safety mechanisms and avoiding accidents. To achieve this, the MobileNet architecture was selected. MobileNet is designed for efficiency on embedded and resource-constrained systems, striking a balance between accuracy and computational cost. Its key innovation lies in the use of depthwise separable convolutions, which replace traditional convolutions with a two-step process: depthwise convolution to apply filters independently to each input channel, followed by a pointwise (1x1) convolution to combine the resulting features. This structure dramatically reduces the number of parameters and computations while maintaining good accuracy, making it highly suitable for real-time detection in industrial environments. For this project, MobileNet was adapted by reducing the number of layers and modifying the final fully connected layer to predict only two classes: “fallen person” and “not fallen person”, instead of the 1000 classes of the original model.

A first training was performed from scratch without transfer learning and a minimal preprocessing on images, with standardization being the only step before the inference. A broad and diverse set of datasets was used for training and evaluation to enhance robustness, including VFP290K, the UR Fall Detection Dataset, the Université de Montréal and Université de Bourgogne datasets and the Universidad de Alcalá dataset. The combination of multiple datasets exposes the model to varied real-world scenarios, making it more resilient to different poses, camera perspectives, and environmental conditions.

5.1.2.4 Unexpected situations detection and localization by UniTO and RULEX

The goal of this component is to detect generic anomalies in the scene. Video frames are divided into four safety areas to learn and detect whether unexpected conditions are occurring in the scene. Using these safety areas, four AI-based detectors are trained to recognize unexpected conditions into the test data. Each detector is based on a Variational AutoEncoder–Generative Adversarial Network (VAE-GAN) approach. The model is trained to reconstruct regular situations through encoding-decoding steps. When anomalous frames occur, the encoder-decoder is not able to reconstruct them, and therefore the quality of the reconstructed image will be poor (i.e., it will not faithfully replicate the input frame). The difference between the input and reconstructed image serves as a score measuring how anomalous the input frame is.

Training of these four detectors took 42.5 hours using an NVIDIA RTX-4070 GPU. Threshold is computed using existing anomaly scoring methods and a set of newly introduced pixel tolerance-based methods (with kernel distance, Gaussian, and quantile tolerance). Hence a total of 53 methods is searched to tune the thresholds. Based on this, the best performing anomaly score function and threshold are selected to be used for all safety areas.

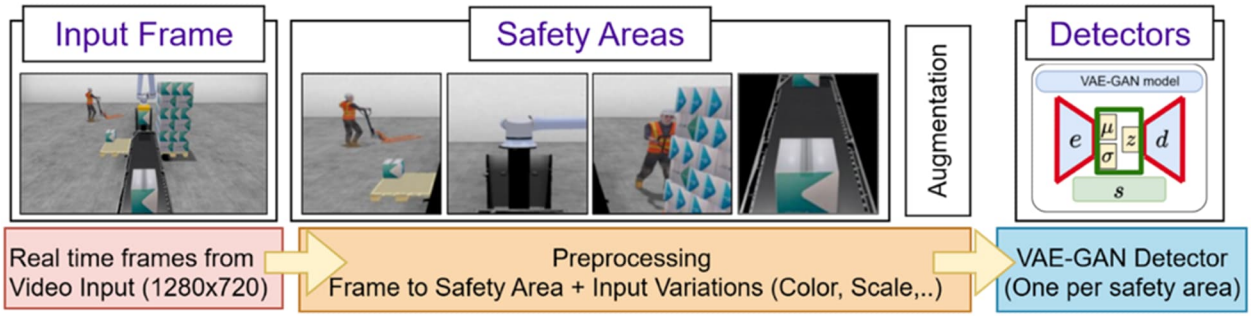


Figure 76 Training Flow

Inference is performed on test data (in this case an unexpected person entering into the scene, but other anomalies could be tested as well). The video data is analysed using trained AI detectors which provides detection score and detection map. This score is normalized with the tuned individual lthresholds (*tau*).

The area is labelled as UNEXPECTED if the normalized score is greater than 1, normal otherwise. Figure 77 shows the unexpected scenario where an unauthorized person enters the scene and moves close to the left pallet: the deviations is captured by the “pallet left” detector and highlighted with a scaled score of 3.09 and safety area is labelled as UNEXPECTED.

Expected Case: Robot is palletizing the boxes on pallets and when one pallet is filled, it is replaced with the empty one

Unexpected case: a human enters the workspace, a box falls, or lighting suddenly changes—clearly highlighted in the saliency maps.

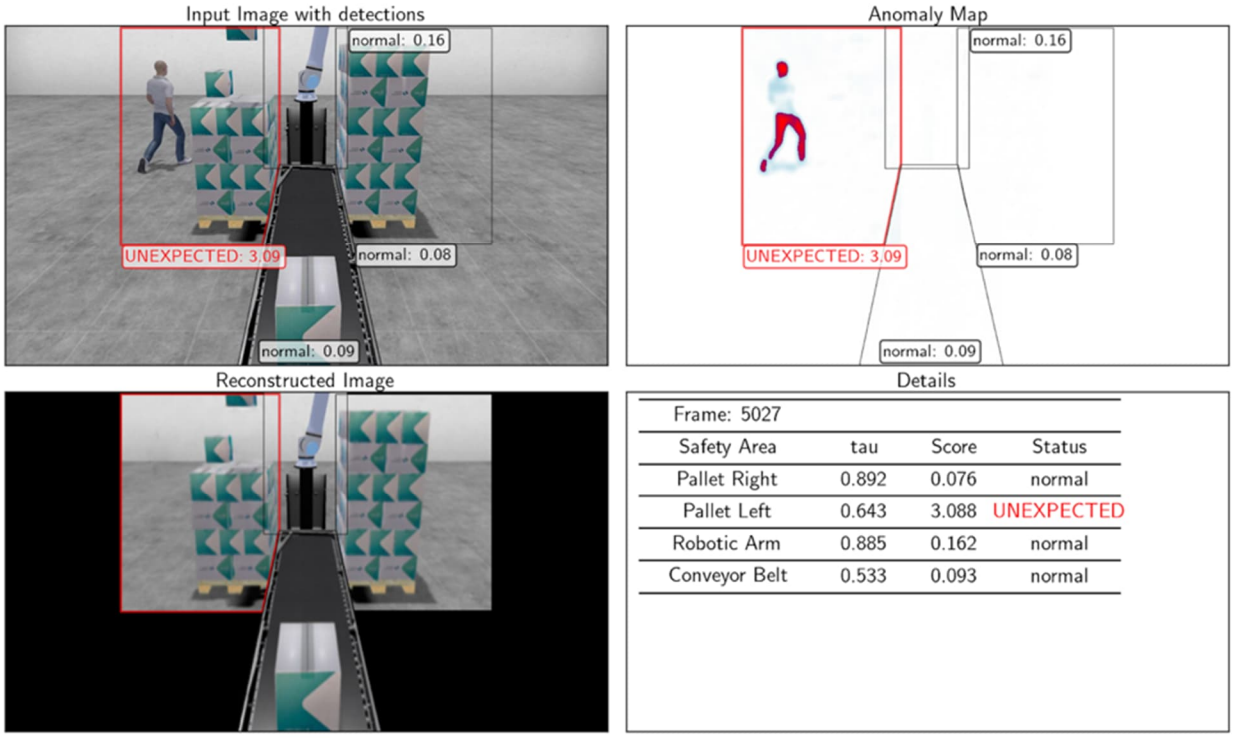


Figure 77 Inference on simulated data with separate detection in safety areas

Overall performance is evaluated with the annotation file generated with the help of segmentation maps. The test video contains 5861 frames becoming 23444 (5861x4) safety areas from which 23211 are detected as normal and 233 as UNEXPECTED. Total incorrect cases contain 70 false positives and 22 false negatives. The overall accuracy of the system is 99.61% while the precision,

recall, and F1-score are 87.4%, 95.1% and 90.9% respectively. A critical analysis is performed for incorrect predictions which have been found to be mostly on edge cases where a minimum count of pixels entered the safety area and therefore the UNEXPECTED class was not correctly recognized. Few other cases were in the robot arm safety area.

5.1.2.5 Deterministic network communications and monitoring by SED

Within the P3-GRA pilot, SED has provided the Time-Sensitive Networking (TSN) infrastructure required to interconnect the different devices. The setup consisted of three PCs employed by UNIGRA for virtual reality simulations, which transmitted messages across the TSN network containing both camera content and control commands. By ensuring the correct configuration of the network—performed through the Central Network Controller, which enables centralized control and monitoring of all network nodes—it was possible to guarantee bounded latency and proper traffic differentiation. This configuration allows critical messages, such as control commands, to be assigned higher priority, ensuring reduced latency and the elimination of message loss. At the same time, it reserves sufficient resources for camera streams, which are less critical but still essential for the overall system functionality.

The ability to prioritize different traffic classes is particularly relevant in robotic applications. In such scenarios, the loss of control messages can lead to critical consequences, and therefore, these messages must always be delivered within a strictly defined time window. TSN technology, when combined with a centralized management tool, provides the reliability and determinism required to meet these demanding constraints, ensuring that both safety and operational efficiency are preserved.

This setup will be integrated into the P3-BEST, enabling a seamless transition from simulation-based development to deployment in real-world robotic environment and allowing complex systems that rely on both high-bandwidth data streams and time-critical commands to operate within defined performance bounds while maintaining reliability.

5.1.2.6 Data fusion in a joint world model by SR

Work on the DistriMuSe data fusion challenge was initiated, focusing on how multi-modal sensor information can be combined into actionable insights for safe and efficient human-robot collaboration. While facing some delay on this topic, the approach has been defined and laid the foundation for development.

The strategy builds on the consortium’s diverse pilot technologies, such as Pumacy’s posture detection from IMU wearables, Emoj’s ergonomic assessment from RGB video, AITEK’s fallen-person detection, and UniTO/RULEX’s anomaly recognition in work environments. Each of these modules generates valuable but heterogeneous data streams—ranging from numerical time-series to image-based classifications. The challenge is to integrate these inputs into a coherent decision-making pipeline that allows the robot to adapt its behavior in real time, or to provide meaningful alerts to the operator through the Human-Machine Interface.

To achieve this, deep learning–based multi-modal fusion architecture is explored. The setup consists of modality-specific encoders—for instance, a convolutional encoder for vision-based detections, and a lightweight transformer for event-level anomaly scores. These encoders extract compact, high-level feature representations, which are then fused into a shared latent space. A decision-making decoder maps this representation to the next robot state, such as adjusting speed, pausing

movement, or issuing a warning. This approach has the advantage of being flexible to new modalities and scalable as additional sensors or partners' algorithms are integrated.

In parallel, the ROS-based data pipeline is developed to connect the fusion module to the robot and HMI. This includes synchronization of inputs, normalization of heterogeneous formats, and logging for traceability. Early experimentation is being carried out in simulation, before moving to on-robot testing.

By combining insights across human posture, ergonomics, event detection, and environmental anomalies, our work will enable the robot to act not only reactively but also proactively—improving safety, ergonomics, and overall system intelligence. This marks the first step towards a robust sensor fusion and decision-making module at the core of DistriMuSe.

5.1.2.7 Digital twinning by UNIGRA

Within the P3-GRA pilot, UNIGRA has developed an advanced simulation framework based on Unity for reproducing complex robotic environments in virtual reality (see Section 5.2). This framework, originally conceived for dataset generation and training of anomaly detection algorithms, will be integrated into the P3-BEST pilot as a digital twin of the collaborative robotic cell. The objective is to provide a real-time, synchronized representation of the state of the physical system, serving as a continuous monitoring and risk detection tool.

The digital twin will be tightly connected to the physical setup through the ROS/ROS 2 middleware, which enables the collection and distribution of real-time data from all relevant components. This includes the operational status and kinematics of the robotic arm, the position and activities of human operators, the state of the conveyor belt, and the dynamic handling of pallets and other work objects. By mirroring this information into the Unity-based environment, the system creates a virtual replica of the workspace, continuously updated at runtime.

The integration of this technology will allow not only a visual reconstruction of the environment, but also the application of advanced analytics on top of the mirrored data. It enables the detection of potentially unsafe situations, such as unexpected operator movements or unsafe proximities between humans and the robot. When such risks are detected in the digital twin, the system can trigger appropriate responses in the real environment, such as slowing down or halting the robot, issuing safety alerts, or logging the event for further analysis.

By transferring the technologies developed in P3-GRA into P3-BEST, the consortium ensures a seamless connection between simulation-driven development and real-world deployment. This approach not only validates the representativeness of the simulated data but also demonstrates the tangible impact of digital twinning as an enabler of safer, more trustworthy collaborative robotics.

5.1.2.8 HMI strategies by UNISOB

The study of Human–Machine Interaction (HMI) strategies focuses on designing and empirically testing communication modalities that can effectively enhance safety, awareness, and operator trust in human–robot collaboration. To this end, UNISOB and RELAB, in collaboration with SR and UNIGRA, have developed a Virtual Reality (VR) simulation environment that replicates the operational conditions of a palletizing robotic arm. The VR environment builds directly on the graphical assets of the digital twin described in the previous section (namely, the 3D model of the robotic arm, the conveyor belt, and the surrounding workspace), reused as the visual foundation of the simulation environment. However, the control logic and movement routines of the robotic arm were edited to meet the specific requirements of the simulation to be provided for conducting HMI

tests with experimental participants. While the original digital twin focuses on real-time synchronization with the physical cell for monitoring and risk detection, the VR environment exploited for user tests is conceived as an interactive simulation platform. This immersive setup allows controlled experiments in which operators are exposed to both ordinary tasks (e.g., fixing a fallen box) and anomalous events (e.g., operator falling in the robot's workspace), within engaged and disengaged areas around the robot. The experimental protocol compares different Human Monitoring-enabled HMI strategies. In the first configuration, called DistriMuSe 1, the robot adapts its behavior to the presence of the operator (for example, slowing down or stopping) without providing any explicit feedback. This approach tests the extent to which behavioral adaptation alone can support safety perception and operator situational awareness. In the second configuration, called DistriMuSe 2, the same adaptive behavior is enriched with explicit communication of operator recognition and status awareness through two advanced channels. The first, also called DistriMuSe 2a, is the Operator Mirroring Ring, a LED-based visual interface mounted on the robot's structure that provides directional feedback to the recognized operator (Figure 78). The second, also called DistriMuSe 2b, is the AR Operator Recognition App, a wearable interface that displays personalized recognition messages through smart glasses (Figure 79). These multimodal communication strategies are designed to make the robot's adaptive decisions transparent to the user, fostering mutual understanding and reducing ambiguity in collaborative contexts. Through the systematic evaluation of user responses across these strategies with VR testing, we will explore how the combination of behavioral and communicative adaptations may contribute to improving safety, workload management, and user experience for operators in collaborative industrial scenarios.

Figure 78 Operator Mirroring Ring. Examples of interaction with the robot in ordinary and anomalous scenarios. The LED-based interface provides immediate visual feedback on operator recognition and the state of collaboration.

Figure 79 AR Operator Recognition App. Feedback through an augmented interface in collaborative scenarios. The AR application displays personalized recognition messages and safety alerts directly through the operator's smart glasses.

5.1.3 Interfacing

Interfaces between the above-described technological building blocks are based on ROS, as commonly used in robotics, and on the REST API, as more generally used.

As an example, the communication protocol between the robot and the HMI, using a REST API, is defined according to the messages as given in Table 21.

Table 21. Communication protocol between the robot and the HMI, using REST APIs.

Definition	Description	Possible Values
`Activity`	Describes the activity within an area	• `NONE`: There is no person present in this area • `NORMAL`: There is a person present in the area, with expected behavior (e.g. switching a pallet, or simply walking by) • `ANOMALY`: There is a person present in the area, with unexpected behavior (e.g., a person is falling)
`RobotMode`	Describes what the current mode is of the robot, given the current area activities	• `FULL_SPEED`: Robot operates at full speed • `LOW_SPEED`: Robot operates at lower speed • `STOP`: Robot will not move
`RobotCommand`	Describes a command to the robot	• `START`: Robot should start • `STOP`: Robot should stop
`RobotStatus`	Describes the current status of the robot	• `RUNNING`: Robot is running • `STOPPED`: Robot is stopped • `STARTING`: Robot is starting up (transitioning from STOPPED to RUNNING) • `STOPPING`: Robot is stopping (transitioning from RUNNING to STOPPED) • `ERROR`: Robot is in error.

Pilot technical KPI measurements

With this demonstrator, we intend to prove the following technical KPI's:

Table 22. P3-BEST technical KPIs

KPI	Validation metric	Status

Imaging radar-based sensing for human motion detection: The evaluation is based on the gained resolution enhancement and the accuracy to detect body part movements	3D and/or polarized radar. Detection of body movements accurate to 90%	Radar modules not integrated in this pilot. Demonstrator will rely only on RGB + IMU fusion.
Positioning of people in indoor environments: Accuracy of estimates of personal indoor trajectories, with current RF + IMU technology is limited by sensor error and drift	Accuracy improvement from 1.5 to 1 m error for 90 % of position estimates in the same environment	KPI assessment will take part in the second cycle.
Late fusion: Ability to track persons temporarily not detectable by sensors, evaluated using Recall and accuracy	Recall: 95% accuracy: 70% correct detection in chosen real-life scenarios after temporal occlusion	KPI assessment will take part in the second cycle.
Explainable AI: User's trust and reliance on the AI system based on the provided explanations	Experts agree with 90% of the AI decisions	Visual evaluation of Unexpected Condition Detector as developed by UniTO and RULEX shows behavior is working as expected. Quantitative analysis results in 99.61% accuracy and 90.9% F1-score.
Activity detection: Develop possibility for AGV and robot cells to detect moving human	Be able to accurately distinguish and avoid moving humans and act upon information on human activity	Human posture/activity detection (PUMACY, Emoj) successfully validated with IMUs and RGB. System distinguishes walking, bending, seated, lying, etc. KPI met in lab conditions.
VR models of human-robot interaction: Generate alerts of possible collisions between the robot, the operator and	Similar to the case using cameras but now further supported with models-on-the-loop.	Validated in Demo 3.2 virtual environment. Collision risks simulated, alerts generated in real time. KPI met in VR, transfer to physical pilot planned in cycle 2.

other elements in the scene.		
Synchronisation and real-time operation: Achieve a deterministic low-latency communication between sensors, processing platforms and actuators.	About 100 us for the most critical traffic up to 100 ms for best-effort. Up to 12 traffic types with up to 4 priority queues and time awareness.	TSN infrastructure tested successfully in P3-GRA. Setup with two TSN nodes, 0.3ns synchronisation offset and three traffic types, each with a different priority. Latency tests carried out in the laboratory, pending integration. KPI achieved in simulation; deployment in P3-BEST scheduled for cycle 2.

To evaluate the success of the integrated technologies, a multi-level assessment methodology will be applied, combining quantitative performance metrics, real-world scenario validation, and expert user feedback. The focus is to ensure the reliability, robustness, and certifiability of the safety system for human-robot interaction in industrial palletizing environments. The methodology includes the components as depicted in the diagram below.

Figure 80: KPI Assessment Methodology

5.1.4 User aspects: stakeholder engagement in pilot development

Concerning the engagement of industrial stakeholders acting as domain experts, for P3-BEST, the same two-stage engagement methodology adopted for P2-REGG will be applied, involving subject matter experts through formative and summative engagement to collect structured stakeholder feedback.

The use cases addressed in this pilot include the design of human–machine interfaces for a palletizing robot that adapts its movements to the operator's presence, and the development of an anomaly detection system for the control room of robotic arms. The expert panel will consist of industrial experts with domain knowledge relevant to these scenarios. For the palletizing case, specialists in human–robot collaboration, industrial safety, and production management, as well as industrial experts in robotics operations and control systems, such as control room engineers and specialists in anomaly detection technologies will be sought. The identification and recruitment of these experts is currently ongoing in collaboration with Smart Robotics.

The engagement will be structured in two stages. The first, defined as formative engagement, will focus on the validation of scenarios and the collection of expert views on risks, opportunities, and adoption barriers. Semi-structured interviews will be supported by design graphics and visual materials illustrating the pilot scenarios. These materials will provide a shared frame of reference, enabling experts to deliver targeted and actionable feedback. The interviews are planned for October 2025, so that the results can directly inform the refinement of scenarios and the subsequent development of the pilot.

The second stage, defined as summative engagement, will take place once the demonstrators have reached a sufficient level of maturity. The same experts will be re-engaged to revisit their initial feedback and compare it with the actual implementations achieved by the project. Demonstration sessions with interactive prototypes will support this reflective assessment, allowing experts to comment on usability, acceptance, and the way the solutions address the risks and opportunities identified during the formative stage.

5.1.5 User-based KPI assessment

In the context of P3-BEST, a structured experimental protocol was developed by UNISOB to evaluate key user-centered indicators, namely safety awareness and mental stress, during human-robot collaboration. The assessment was conducted in a virtual reality (VR) simulation environment developed by RELAB and UNISOB, in collaboration with UNIGRA and SR. This environment realistically replicates the operational scenario of a palletizing robotic arm designed by SR, allowing participants to experience tasks from the operator's point of view and enabling the exploration of human-machine interface (HMI) strategies under realistic yet controlled conditions. By combining immersive simulation with validated measurement methods, the study aims to gather robust insights into user experience, even in the absence of a physical robotic system.

As outlined in D1.2 and reiterated here for clarity, mental stress is defined as a psychophysiological state that can impair perception, decision-making, and action. In industrial contexts, it can stem from cognitive overload, emotional fatigue, or disengagement, compromising both performance and perceived safety. In this study, mental stress was assessed using the **NASA Task Load Index (NASA-TLX)**, a widely used subjective tool that captures perceived workload across six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Participants rated each dimension on a Likert scale, and the average provided an overall workload estimate.

To better understand the emotional perspective, the **Self-Assessment Manikin (SAM)** scale is also administered, allowing for the evaluation of perceived arousal, an emotional correlate often linked to stress levels in demanding tasks. Namely, the Self-Assessment Manikin (SAM) is a non-verbal pictorial tool used to evaluate emotional responses along three dimensions: Valence (pleasantness), Arousal (intensity), and Dominance (sense of control). To further capture participants' emotional responses, the Self-Assessment Manikin (SAM) scale was also administered after each scenario. SAM is a non-verbal pictorial tool used to evaluate emotional experience along three dimensions: Valence (pleasantness), Arousal (intensity), and Dominance (sense of control).

Safety awareness, in turn, refers to the operator's awareness of safety-critical elements in the environment, such as anticipating the robot's movements or recognizing potential hazards. This was assessed using the **Situation Awareness Rating Technique (SART)**, a validated tool composed of three subscales: Demand (complexity and pace of the situation), Supply (available cognitive resources), and Understanding (operator's comprehension of the situation). Higher SART scores indicate a greater sense of awareness and control. To deepen the analysis, participants also answered a set of ex-post ad-hoc questions designed to probe their mental model of the robot, specifically its behavior, and predictability in the shared task context.

The evaluation protocol was structured to immerse participants in realistic human-robot collaboration scenarios. Wearing a Meta Quest 3 headset, each participant completed four short interactive scenarios, each lasting approximately two minutes. The tasks were situated in two areas of the workspace: the engaged area, where the robot actively loaded boxes onto a pallet, and the disengaged area, where a second pallet was present but not involved in robotic operations. In both areas, participants carried out ordinary actions (e.g., inspecting pallets or repositioning fallen boxes) and anomalous actions (e.g., simulating a fall or injury near the robot). These variations yielded four distinct conditions, which were presented in counterbalanced order across four experimental groups to mitigate order effects such as learning, fatigue, or emotional habituation.

At the end of the interactive session, participants completed a final questionnaire assessing perceived workload (NASA-TLX) and emotional responses (SAM) for each scenario. They also completed the SART scale and responded to targeted questions concerning their mental model of the robot's behavior and goals. Before taking part in the experiment, all volunteers were informed both verbally and in writing by the experimenter, signed a written informed consent form, and completed a pre-experiment questionnaire to collect demographic information, prior experience with VR and robotics, and individual traits relevant to sample characterization.

Table 23 Aggregated results of the subjective metrics collected during the assessment.

User-based KPI	Total
Perceived workload	Total score: M = 2.54 Mental Demand: M = 2.98 (range 1-7)
Situation awareness	M = 5.94 (range 1-7)
Other subjective metrics	**Total**
Perceived emotions	Valence: M = 2.89, SD = 1.04 Arousal: M = 2.75; SD = 0.95

	(range 1-5)
Robot model comprehension	M = 3.35, SD = 1.10 (range 1-4)

A total of 48 participants volunteered for the experiments. Preliminary results offer valuable insights into how participants perceived the robotic system in the baseline condition. The average situation awareness score (M = 5.94) suggests that participants had a high level of clarity and understanding of the task environment. This aligns with the scenario's design: since the robot's behavior remained fixed and independent of user actions, there was little variability to interpret. The task required minimal continuous adjustment or inference, which likely contributed to the high reported awareness.

As for perceived emotions, average scores for valence and arousal reflect a moderate level of emotional engagement. These values should not be interpreted as low, but rather as indicative of a neutral and predictable context: the robot did not exhibit any sudden changes or emotionally salient behaviors, and the overall experience was stable. This emotional profile fits the role of the baseline scenario as a controlled reference condition, providing a benchmark for future comparisons involving more adaptive systems.

Perceived workload was also low to moderate (M = 2.54), with slightly higher values for Mental Demand (M = 2.98). The absence of pressure, time constraints, or complex decision-making tasks reflects the simplicity of the baseline, which was designed to simulate realistic actions without cognitive overload. Again, this condition is not meant to be adopted in real-world settings but serves as a comparative reference for upcoming evaluations.

Importantly, the baseline scenario does not represent a deployable industrial solution. In its current form, the robot's speed exceeds what would be acceptable in a fence-free setting. While efficient from a production standpoint, this level of speed would not meet safety requirements without physical barriers. The goal of this user-centred study is to explore alternative interaction models based on human monitoring, capable of adapting robot speed according to operator proximity and behavior. These solutions could enable safe, fence-free configurations, optimizing space and improving workflow continuity in contexts where traditional safety mechanisms would require a full stop of robotic activity.

The first evaluation of a human monitoring-based solution without any multimodal HMI is planned for September 2025 and will allow for a direct comparison with the current reference data. This upcoming study will further refine user-centered parameters and support the development of safer, more adaptive human-robot interaction models.

5.1.6 User aspects: Gender/age issues and ethical concerns in P3-BEST

The ethical analysis addresses all three pilots within the robotics domain of DistriMuSe (UC3), due to the strong conceptual and technical commonalities they share. P3-BEST focuses on industrial plants equipped with palletizer robots and leverages human monitoring to understand the position and condition of operators, allowing the robots to adapt their behavior in real time to improve both operator safety and overall system performance. P3-GRA is designed to detect risky situations involving human operators in similar environments, using distributed cameras and virtual reality-generated datasets to train detection algorithms. P3-SON, by contrast, targets scenarios involving mobile unloading robots, aiming to monitor human presence and support industrial plant controllers in identifying hazardous conditions on the floor, while also enabling safe interaction between humans and robots through movement prediction.

Across the three pilots, the primary users are industrial plant controllers responsible for monitoring system status and responding to alarms and industrial plant operators, who work in close proximity to the robots that rely on human monitoring systems to operate safely and effectively. All three solutions aim to enhance operator safety and situational awareness through AI-based human monitoring technologies. They are all deployed in fenceless industrial settings where humans and robots share the same workspace, thus requiring advanced coordination mechanisms and robust risk mitigation strategies. The joint focus on the recognition of human operator presence and movements, as well as anomalous or risky situations through artificial intelligence further reinforces the coherence of the ethical reflection across pilots. For this reason, conducting a single integrated focus group enabled a unified approach and encouraged cross-pollination between development teams, allowing each pilot to benefit from complementary perspectives.

The focus group involved nine participants, all directly involved in the management or development of the three pilots. Their roles included algorithm developers, HMI designers, R&D project managers, and a strategic lead with executive responsibilities. The group brought together expertise in explainable AI, computer vision, sensor fusion, robotics control and hardware, safety regulations, user-centred design and user experience, applied machine learning, and virtual reality. This diverse combination of operational, design, and strategic perspectives enabled a well-rounded and multi-level ethical discussion.

The focus group was organized by the partner responsible for coordinating the ethical reflection activities (UNISOB), based on the guiding questions provided by the DistriMuSe Ethics Exercise Tool. The session was conducted online on July 29, 2025 and involved nine participants, all directly engaged in the development or coordination of the three pilots.

Participants were fully informed about the processing of shared data and each signed an informed consent form prior to participation. The discussion was structured using an online collaboration tool (MIRO), which supported a two-phase process aligned with the five key dimensions of the exercise: background, anticipation, reflectivity, inclusiveness, and responsiveness. In the first phase, individual contributions were collected asynchronously to avoid mutual influence; this was followed by a synchronous roundtable discussion to deepen and elaborate the shared reflections.

Following the session, the partner in charge synthesized and organized the results and circulated the draft responses among pilot partners for final validation. The completed input was then submitted to the WP7 team in line with the project's ethical reporting plan.

Detailed results of the ethics exercise will be included in deliverable D7.7 (WP7).

5.1.7 P3-BEST transition into cycle 2: takeaways and feedback

Highlights of cycle 1. In P3-GRA (the virtual twin of P3-BEST) we validated the core building blocks needed for an intelligent safety system. IMU-based posture monitoring showed stable BLE streaming (>99.8% packets) and an end-to-end pipeline proven with an LSTM baseline. Vision modules matured: ergonomic scoring from RGB keypoints proved robust to viewpoint changes; fallen-person detection ran efficiently on MobileNet; and environment anomaly detectors achieved high offline accuracy/F1 with clear saliency maps. UNIGRA's Unity digital twin enabled fast iteration and safe exploration of edge cases, while SED's TSN setup demonstrated determinism and priority handling for mixed video/control traffic. Together, these results de-risked the sensing and networking assumptions behind our fusion approach.

Lowlights of cycle 1. Most evaluation occurred in simulation; on-robot, fence-free validation in Best is still outstanding. Multi-modal fusion remained largely architectural (encoders + shared latent) with limited end-to-end testing across all modalities at once. Practical wrinkles emerged: cross-sensor

time alignment, camera–IMU calibration, and managing false positives at safety-area boundaries. The fusion work started later than planned, reducing time for ablation studies and HMI-in-the-loop trials.

Intended changes for cycle 2. We will shift focus to integration and testing on the physical P3-BEST cell in Best. Concretely:

1. **System integration:** harden the ROS/ROS 2 interfaces with a common timebase (PTP over TSN), unified message schemas (pose/posture/anomaly with confidences), and per-module health/status topics.
2. **Fusion & safety logic:** implement the multi-modal encoder–fusion–decoder end-to-end.
3. **HMI-in-the-loop:** bring the Operator Mirroring Ring/AR feedback into the physical loop and assess how feedback changes operator workload/awareness during real interventions.
4. **Evaluation protocol:** progress from log-replay → "shadow mode" (advisory only) → gated actuation at reduced speeds, with KPI tracking (latency, recall after occlusion, ergonomic-risk reaction time).
5. **Data campaign:** collect synchronized RGB/IMU/robot data in Best to fine-tune fusion thresholds and reduce edge-case false positives.

These changes turn cycle-1 technology tests into a deployable, measurable safety function on the real system.

5.2 P3-GRA evaluation cycle 1

5.2.1 Final pilot set-up

The P3-GRA pilot, coordinated by UNIGRA, investigates the use of virtual reality simulations combined with digital twin architectures to enhance human safety in robotic environments. The pilot aims to identify operational risks, evaluate human-machine interfaces (HMI), train AI models, and optimize sensor configurations through immersive simulation of the scenes. This is achieved via a collaborative effort involving multiple partners, each contributing to a specific layer of the system:

- UNIGRA is responsible for the overall integration of the simulation environment, including the development of the Unity-based VR scenes and the orchestration of the distributed simulation architecture.
- UNITO and RULEX jointly develop the AI-based anomaly detection system, which relies on multi-camera inputs (including segmented views) to recognize abnormal or unsafe behavior around the robot. RULEX also contributes its explainable AI framework to support transparent and traceable decision-making.
- AITEK and EMOJ implement vision-based analysis of human movement, focusing on the operator's posture and behaviour to identify dangerous or inappropriate actions during interaction with the robot.
- UNISOB and RELAB lead the evaluation of user aspects and HMIs, assessing usability, acceptability, and human factors in the simulated environment. This includes user-based KPIs, qualitative feedback, and ethical considerations.
- SED oversees ensuring communication efficiency and latency control within the distributed architecture, optimizing data flows between the simulation components, sensors, and analytical modules.

- RESILTECH contributes to the safety and risk analysis of the digital infrastructure, identifying potential vulnerabilities and failure modes that could compromise user safety or system integrity.

5.2.2 Pilot evaluation execution and protocol details

Updated simulation architecture for dataset generation

During the first evaluation cycle, a significant change was introduced in the simulation architecture, compared to the initial plans outlined in D5.1. Specifically, the dataset generation workflow was redesigned as a two-stage pipeline (Fig. 5.2.2), separating the simulation and rendering processes for greater control and flexibility.

- **Stage 1: Simulation and scene state recording**. A modular and distributed simulation system is used, combining:
    - Unity to define and animate the physical environment (humans, objects, scene dynamics).
    - URSim to simulate the robotic arm's behaviour and interactions, using the same interface as the physical robot.
    - ROS2 as middleware to handle all inter-node communications and timing control.

In this stage, all simulation components run in parallel, and the entire state of the scene is stored in a ROSBAG file, sampled at 60 Hz (with potential to support higher frequencies).

- **Stage 2: Rendering and camera output generation.** A standalone Unity-based application replays the ROSBAG to reconstruct the scene. A configuration file in JSON format defines the camera setup: number, type (RGB, segmented), resolution, frame rate, image compression, and position/orientation. Based on this configuration, the system generates synchronized video sequences and frame-level image sets, which are used for training and validating AI models.

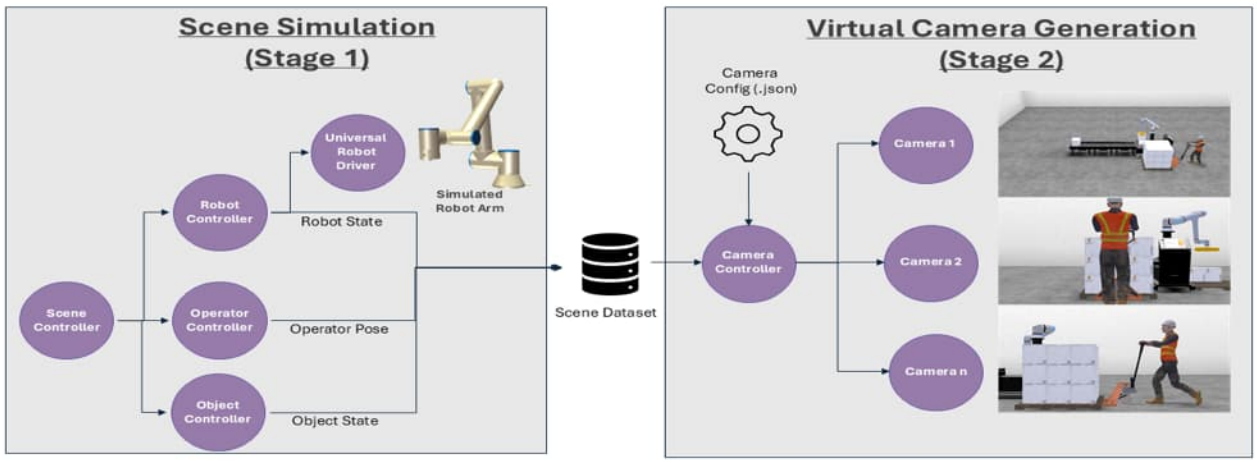


Figure 81 Updated two-stage data generation architecture implemented in the P3-GRA pilot. In the first stage, a distributed simulation system composed of Unity, ROS2, and URSim generates a synchronized ROSBAG containing the full state of the scene at 60 Hz. In the second stage, a Unity-based rendering executable uses this ROSBAG and a JSON camera configuration file to generate synchronized video and image sequences

This new architecture provides multiple benefits. It ensures perfect synchronization across all cameras, a critical requirement for training anomaly detection algorithms based on paired RGB and segmentation views. This updated architecture allows the creation of a virtually unlimited number of camera perspectives for the same scene, increasing dataset diversity without re-simulating the environment. Finally, this architecture makes it possible to extend existing datasets with new sensor setups post hoc, supporting future AI models and unforeseen research use cases.

The main trade-off is that real-time digital twin applications will need a parallel architecture capable of generating live data and rendering on the fly, requiring architectural adjustments compared to this dataset-focused offline workflow.

This updated simulation strategy was validated during the first implementation cycle (D5.2) and is being adopted as the baseline for training AI models, conducting HMI evaluations, and ensuring reproducibility of scenario generation within P3-GRA.

In addition to the simulation and rendering pipeline, a set of auxiliary Python scripts has been developed to facilitate dataset handling and visualization. These scripts provide several useful functions for dataset users, including:

- Video generation from raw image sequences, enabling the quick creation of playable demonstrations of the simulated scenes.
- Video composition, allowing multiple camera views to be combined into a single frame for comparative analysis or presentation purposes.
- Image analysis utilities, such as tools for inspecting and verifying colour segmentation in generated images.

These tools have proven especially useful during the first cycle, both for internal validation of dataset quality (e.g., checking segmentation consistency) and for the preparation of presentation videos to showcase simulation results to partners and stakeholders. By making these scripts available

alongside the datasets, the pilot ensures that other users can easily visualize, validate, and reuse the data in a more efficient and user-friendly manner.

Proofs of concept and integration of WP technologies

- From WP1 (Requirements and Use Cases): The scenario selection and definition were based on risk analysis use cases derived from industrial environments involving close collaboration between operators and robotic arms. These scenarios were formalized into simulation scripts and user interaction patterns to be tested in virtual environments.
- From WP2 (Sensor and communication technologies):
    - AITEK's operator movement tracking system was adapted and validated using synthetic camera views. This allowed testing their algorithms without requiring a physical deployment.
    - SED contributed with a middleware layer for communication orchestration, helping coordinate the different modules (e.g., Unity, ROS2, data recorders) while preserving synchronization constraints and bounded latency. Simulated data was used to assess the amount and typology of data expected for the physical scenario.
- From WP3 (AI, fusion, and analytics):
    - UNITO and RULEX anomaly detection system was trained and evaluated using synthetic datasets generated in the pilot.
    - RESILTECH initiated a cybersecurity risk analysis, focusing on identifying possible attack vectors in the simulation-to-decision-making pipeline.
    - EMOJ ergonomic evaluation system was trained using synthetic datasets and offers a system that is independent of camera position and calibration.
- From WP4 (Distributed architecture and platform):
    - UNIGRA led the development of a modular, scalable architecture integrating Unity simulation, URSim for robot simulation, and ROS2 as the messaging backbone.

The system was packaged to support deployment across distributed nodes, allowing parallel execution and decoupled rendering.

5.2.3 Pilot technical KPI measurements

The P3-GRA pilot integrates advanced sensing, simulation, AI, and network technologies. To assess its technical performance, a set of dedicated KPIs was defined in Deliverable D5.1 and the project proposal as indicated in Table 5.2.3. These KPIs target explainability of AI, network synchronization accuracy, and real-time traffic handling in time-sensitive environments.

Table 24 Technical KPIs for P3-GRA

KPI	Validation metrics	Status
2.6. Explainable AI: Degree of comprehension between human experts and AI system decisions.	Human experts agree 90% of the AI decisions	
3.4.1. Network synchronisation error for time-critical operations	Network k-wide synchronisation error ≤100µs (for most critical traffic)	Achieved. Synchronization offset: 0.3ns. Synchronization jitter: 0.068ns. Setup with two TSN nodes.

3.4.2. Capability for TSN traffic handling and shaping based on type and priority.	Up to 12 traffic types with up to 4 priority queues and time awareness	Three traffic types, each with a different priority. Four priority queues in total, including the synchronization maximum priority queue.
3.4.3. Network latency variation (jitter) for deterministic TSN flows.	Jitter between 400 ns and 50 µs (depending on flow/policy)	Tests carried out in the laboratory in which the target has been met. Integration in progress.

5.2.4 User aspects: stakeholder engagement in pilot development

As described in Deliverable D5.1, the primary intended users of the simulation outputs within the P3-GRA pilot are:

- Developers of AI algorithms (e.g., for anomaly detection or behaviour prediction).
- Engineers responsible for the design and configuration of collaborative robotic environments.
- Human operators are expected to interact with robots in industrial scenarios.

During the first cycle of the pilot, UNIGRA has actively engaged the first two user groups (AI developers and robotic system designers) in the iterative design, validation, and refinement of the simulation datasets. All P3-GRA partners had continuous and direct access to preliminary simulation outputs, enabling early-stage feedback regarding:

- The content and structure of scene states and ROSBAG logs.
- The number, quality, synchronization, and format of the multi-camera image sequences.
- The realism and relevance of the simulated scenarios.

UNITO and RULEX played a key role in defining the required dataset structure for their anomaly detection modules. Their input guided the development of the two-stage data generation pipeline, and they also participated in validating the output to ensure compatibility with their machine learning workflows.

Furthermore, all P3-GRA partners contributed to the collaborative definition of the simulation elements and use cases (e.g., robot models, human behaviour, object interactions, number and models of cameras), with the goal of aligning the virtual pilot with the physical implementation planned for P3-BEST in the second cycle.

In parallel, UNISOB and RELAB focused on the third user group: potential human operators. They developed immersive environments based on the same simulation scenes produced by UNIGRA, allowing operators to experience and navigate virtual robotic scenarios. These environments were used to evaluate different HMI configurations, helping anticipate usability and ergonomic factors relevant to future deployment in P3-BEST. Through their engagement with potential end-users, RELAB contributed early-stage insights on human interaction and cognitive workload in collaborative robotic settings.

While the full-scale involvement of real operators will be expanded in the second evaluation cycle, the first cycle successfully ensured that both technical and human-centric stakeholders were involved in shaping the pilot outcomes.

5.2.5 User-based KPI assessment

As highlighted in Section 4.2.4 of D5.1, one of the key user-based KPIs for the P3-GRA pilot concerns the representativeness of the simulation-generated data in relation to real-world use cases and safety-critical situations.

During the first cycle, the evaluation of this KPI has started by systematically defining and implementing simulation scenarios aligned with the expected physical setup in P3-BEST. These scenarios include both normal operations and critical situations that pose safety risks to the operator. However, full validation of data representativeness will only be possible during the second cycle, once the physical environment is operational. This will allow measuring how well AI models trained with synthetic data generalise to real-world sensor inputs.

To support this process, the following data summarises the dataset development status at the end of the first cycle:

- 9 scenes have already been implemented (2 normal operations and 7 critical situations).
- 8 cameras implemented for each scene (4 virtual cameras and 4 with the segmentation of the previous ones).
- ~200 minutes of video have been generated.
- 54GB of video data and 45GB of object state data.

5.2.6 User aspects: Gender/age issues and ethical concerns in P3-GRA

P3-BEST, P3-GRA, and P3-SON jointly carried out the ethical reflection, as detailed in section 5.1.6 "User aspects: Gender/age issues and ethical concerns in P3-BEST", to which the reader is referred.

5.2.7 P3-GRA transition into cycle 2: takeaways and feedback

General summary of the first cycle

The first evaluation cycle of the P3-GRA pilot has successfully demonstrated the feasibility and initial effectiveness of using simulation-based data generation as a foundation for AI model development and HMI evaluation in human-robot collaborative environments.

A significant milestone was the consolidation of a robust, modular simulation architecture capable of reproducing complex industrial scenarios with synchronized multi-camera outputs and semantic annotations. The introduction of a two-stage data generation pipeline, separating simulation (state capture) from rendering (image/video generation), proved critical for ensuring dataset consistency and flexibility. This architecture has become the core of the pilot’s data production strategy and will continue to evolve during cycle 2.

Thirteen simulation scenes were implemented in this phase, many of which represent safety-critical events such as unintended operator movements, occluded robots, or unexpected human presence. These scenes were selected and refined through close collaboration with all P3-GRA partners to ensure their relevance and transferability to the physical pilot (P3-BEST) scheduled for the next phase.

The AI partners (UNITO and RULEX) began training and testing their anomaly detection pipelines using the synthetic data, while RELAB and UNISOB developed immersive environments for early HMI evaluation. Additionally, internal feedback loops allowed continuous validation and correction of scene logic, object placement, and sensor configuration.

Overall, the first cycle laid a solid technical and collaborative foundation for the transition to physical system integration in cycle 2, where the key focus will shift to validating generalisability, usability, and real-world applicability of the components tested in simulation.

Throughout the first cycle, the simulation datasets and tools developed within P3-GRA were actively shared with all involved partners, enabling an early and meaningful feedback process. This internal feedback was crucial to ensure that the synthetic data met the technical requirements of the different subsystems and could be used reliably for downstream tasks such as AI training, human behaviour analysis, and HMI prototyping. This collaborative feedback loop has ensured not only technical robustness, but also early-stage validation from the perspective of diverse user groups. Their contributions will directly inform the creation of new scenes, refinement of avatars and behaviours, and extension of the HMI evaluation protocols in the upcoming cycle.

Technical takeaways and planned updates

The first cycle of the P3-GRA pilot has yielded several important technical lessons that will guide the evolution of the system in the second cycle. Among the most significant is the validation of the two-stage data generation pipeline, which has proven to be both scalable and adaptable to different user requirements. The decoupling of simulation and rendering has facilitated reproducibility, scene reusability, and rapid dataset iteration, making it easier to support diverse AI development workflows.

A particularly valuable outcome of the cycle has been the need to formally define each simulation scene in detail, including task flow, human and robot behaviour, object configuration, and environmental variables. This process has forced the consortium to critically assess and anticipate key aspects of the physical implementation that will take place in P3-BEST. Specifically, it has led to early decisions regarding:

- Which operator actions are considered safety-relevant or potentially hazardous.
- The positioning and layout of physical components in the workspace.
- The number, type, and placement of cameras and sensors required to capture relevant observations.
- The lighting conditions and their effect on visibility and detection performance.

This exercise has significantly improved the definition and planning of the physical pilot environment, ensuring that the virtual and real setups are well-aligned. It also provides a concrete foundation for evaluating transferability from synthetic data to real-world conditions in the next cycle.

Additionally, partners identified the need to enhance the realism of operator behaviour, refine object animations and transitions, and simulate more unexpected or edge-case situations, particularly those that could cause anomalies or safety concerns. These improvements are now planned as part of the simulation updates for cycle 2.

5.3 P3-SON evaluation cycle 1

5.3.1 Final pilot set-up

In Demo 3.3, the partners develop and test a dynamic factory where mobile robots work safely alongside humans. The following scenarios were investigated in the laboratory environment of Prodrive in Son:

- **Person Recognition**: When the mobile robot is unloading a storage bin and a person enters the workspace of the mobile robot, the robot recognises the person and limits its movement in their direction to ensure their safety. When the person leaves the workspace, the mobile robot continues without movement limitations. When another mobile robot enters the workspace, no movements are limited from a safety perspective.

- **Monitoring System**: When the mobile robot moves to an unloading point and unloads, the actions can be monitored on an HMI by an operator. If the safety of the mobile robots is triggered (e.g., a person enters the robot's workspace), it will be visualised. This way, the operator knows what is happening on the floor.

The high-level architecture is illustrated below. The changes have been highlighted in circles.

The HMI Mobile Robot scenario has been prolonged for the second use case cycle because of small delays in development. It is expected that this part will be tested in the second use case cycle (see green circle in high level architecture). The orange circle highlights the components belonging to Trilitec and the University of Bremen, where the first use case cycle was carried out in Bremen's laboratory. The remaining parts of the architecture represent the mobile robots, for which the first use case cycle was conducted in Prodrive's laboratory environment in Son. The remainder of chapter 5.3. P3 – SON will be split into two parts, the use case cycle in the laboratory in Bremen and the AMR within Son.

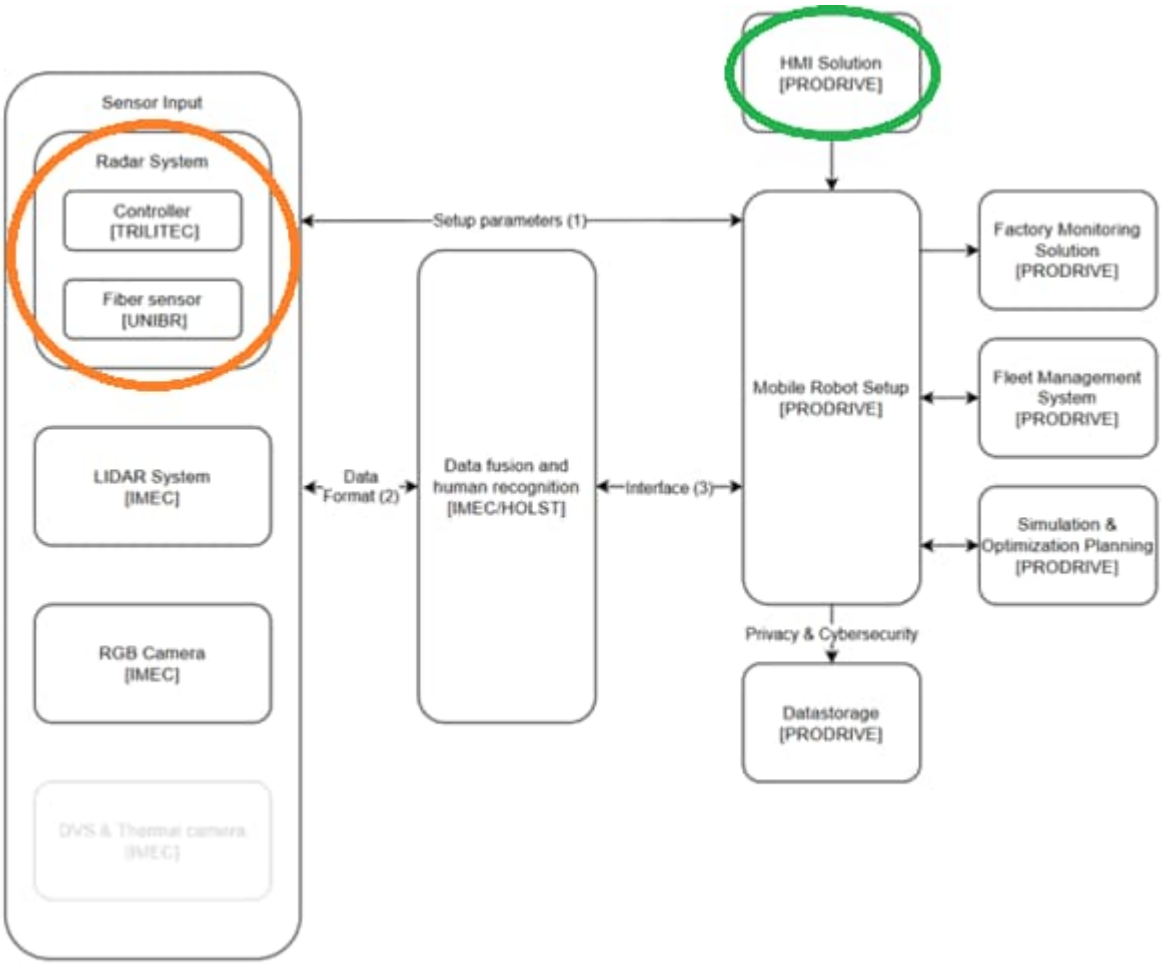


Figure 82 P3-SON Architecture

5.3.2 Pilot evaluation execution and protocol details

5.3.2.1. P3-SON Evaluation cycle 1 in Son

In demonstrator 3.3. the partners expect to collect data with the sensor box of IMEC/HOLST with 5 different inputs. As shown in Figure 83, the sensor box will be mounted on the AGV. There are 5 different inputs, namely radar, LIDAR, RGB, thermal and DVS camera. In the first pilot cycle, the use

case investigates the described components for an AMR driving in reverse, focusing on whether it can detect and respond appropriately when someone passes by or is nearby. The human recognition part is done by data fusion in combination with machine learning. The outcome of this model will then send information to the AMR which behaves accordingly.

Figure 83 P3-SON robots in the factory floor

In this demonstrator the following technologies were tested:

Technologies collected and tested from WP3 (Imec & Holst)

An important component of this pilot is robot environment perception module developed jointly by imec and Holst. We will deploy human detection/tracking/prediction algorithms based cooperative sensor fusion. Imec and HOLST will investigate which combination of sensors yields the best average precision of object classification objects. Furthermore, imec will focus on improvements on its algorithms based on fusion of RGB camera and radar, as well as other combinations with thermal camera and lidar. Besides the fusion of RGB and lidar HOLST will also explore fusion of multiple sensors like DVS camera, radar, and/or RGB for Edge-AI platforms with stringent compute and performance requirements.

Technologies collected and tested from WP4 (PRODRIVE)

The key technology used from WP4 is related to the development of a reliable communication networks for supporting the exchange of data between local sensors and centralized monitoring and control systems. In cycle 1 the main focus is related to:

- Application performance monitoring, centralized platform to monitor the performance of the different applications that control the automated factory.
- Factory monitoring, centralized platform to monitor the mobile robots and other automation. The architecture is shown in Figure 84 and the application itself in Figure 85

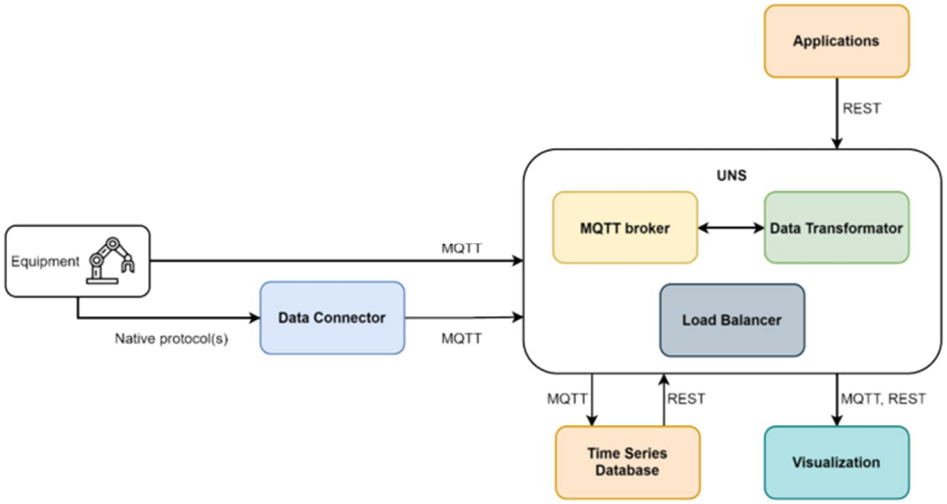


Figure 84 Architecture factory monitoring

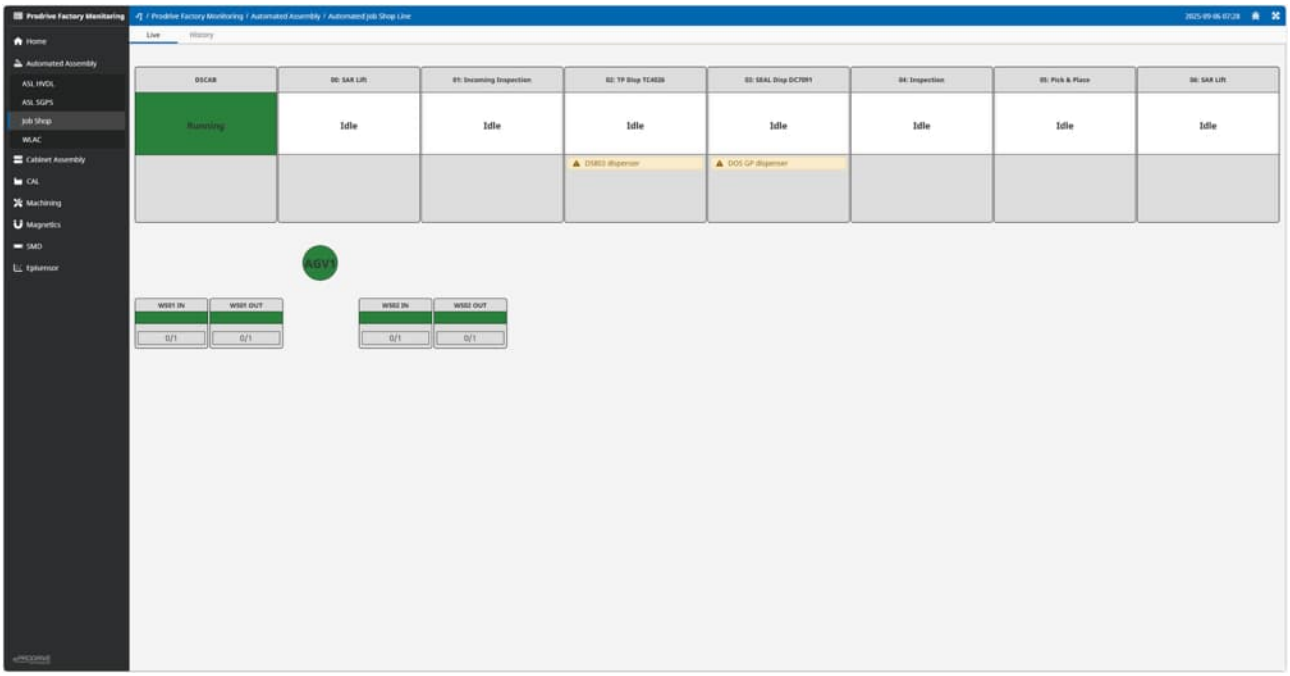


Figure 85 Factory Monitoring Application

5.3.2.2. Pilot Evaluation Execution for P3-SON in Bremen

In this pilot P3 the main target of the UBREMEN and TRILITEC cluster is to bring in an efficient radar that has enhanced sensing capabilities. In this aspect, TRILITEC plays the vital role of developing the radar sensor and UBREMEN develops smart sensing concepts with dielectric waveguides like re-location of antenna and multi spot sensing. In the first cycle for UBREMEN the focus was primarily given to the development and testing of individual components. This includes dielectric waveguide developments, interconnects to the radar system, lens antennas, etc. The results of the individual components are published via IEEE conference proceedings and are detailed in the periodic reports. After the development of these components a validation setup was made in the University of Bremen's laboratory. The setup included a single channel D-band radar unit from TRILITEC and the interconnects + dielectric waveguide + lens antenna setup from UBREMEN. The idea of the test is to check the differences in the performance of the radar unit under two circumstances i) when a

standard horn antenna is used with the radar and ii) when the antenna is relocated with an interconnect to a dielectric waveguide (DWG) and ultimately the lens antenna. For both cases, the corner reflector was used as target. The picture of the evaluation setup can be seen below. In Setup A, the radar unit is connected with a horn antenna that is placed close to the radar and peeks only in the boresight direction from the antenna. However, in the Setup B the same radar unit is connected to a launcher that is guiding the mode into the dielectric waveguide cable which is then fed into the lens antenna. By this way the antenna from the radar is simply relocated to a different place.

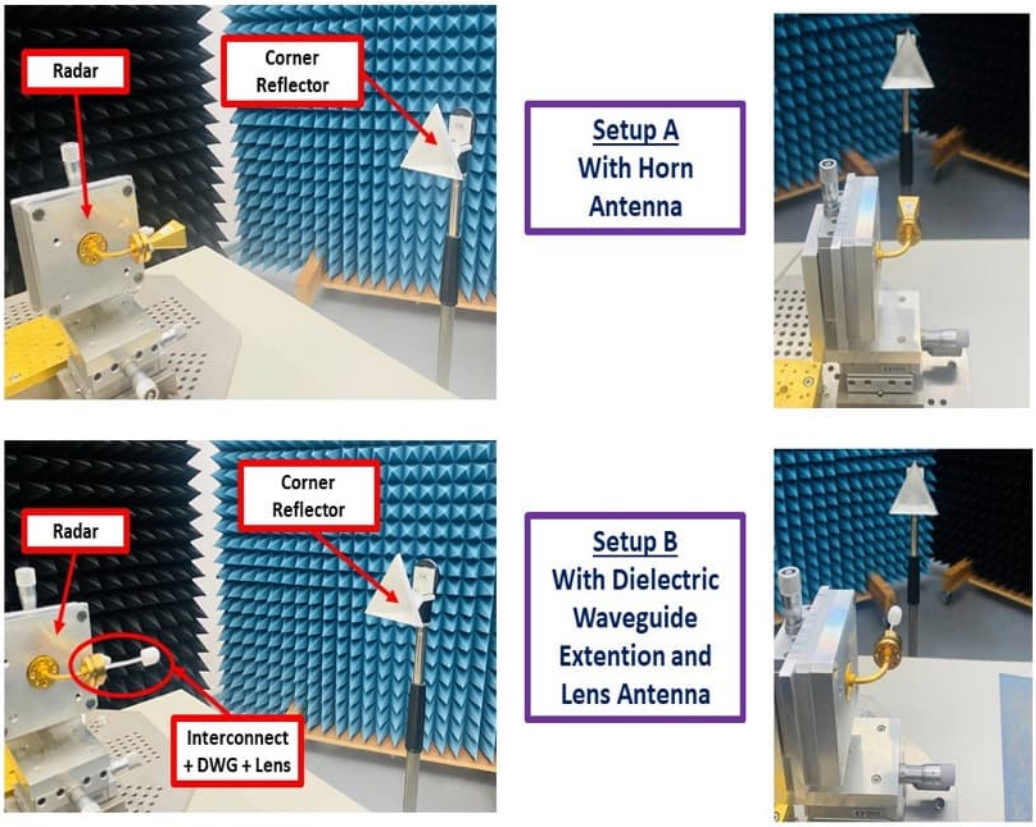


Figure 86 P3-SON set-up of devices

Since the radar currently being developed for this use case will not have a horn antenna but rather an FMCW IC with integrated antennas, Trilitec tested the system's performance with various dielectric (on-chip) lens configurations. This should provide initial insights into the potential suitability

of a dielectric flexible waveguide as an on-chip flange and, in perspective, comparative values for future experiments with the implemented flexible waveguide technology.

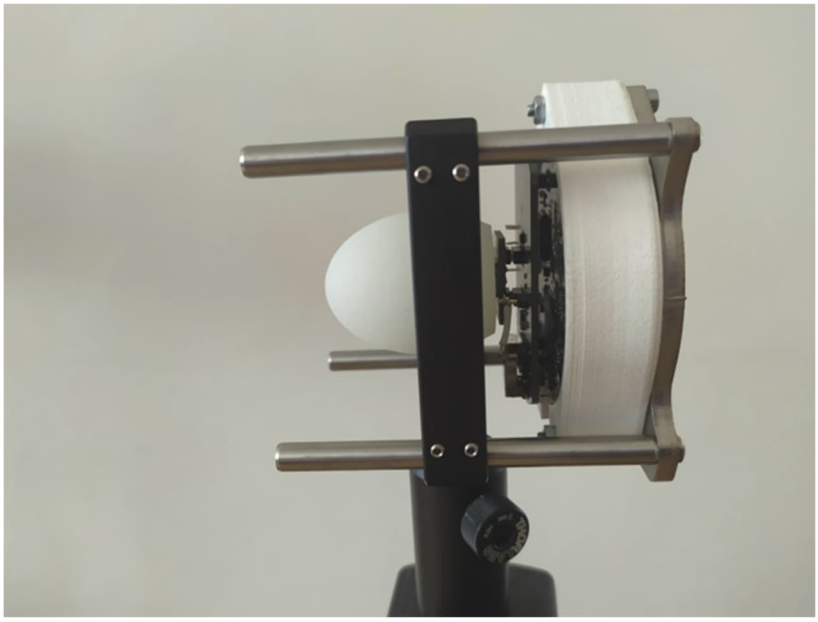


Figure 87 120 GHz radar with elliptical lens flange

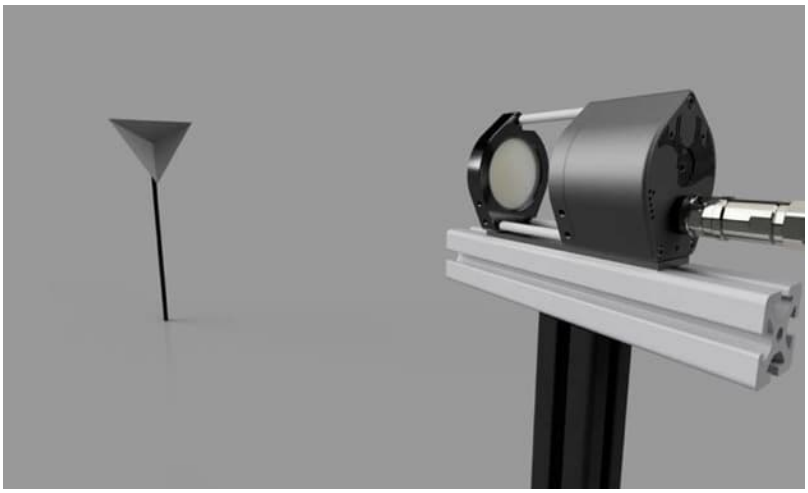


Figure 88 Rendering of test setup for measuring the radar performance with different lenses

Table 25 Measured directivity and beam divergence of different (on chip) lenses.

Lens	directivity in dB	beam divergence in deg.
Far-field lens ⌀ = 50 mm	34,8	2,2
Elliptical lens ⌀ = 30 mm	25,2	4,2
Hyper-hemispherical lens ⌀ = 20 mm	18,9	8,4
Hyper-hemispherical lens ⌀ = 10 mm	12	-

For this purpose, the dielectric lenses listed in Table 25 were mounted on the radar and their directivity was determined using a corner reflector (Figure 87, Figure 88), and the range of each configuration was measured as exemplified in Table 25.

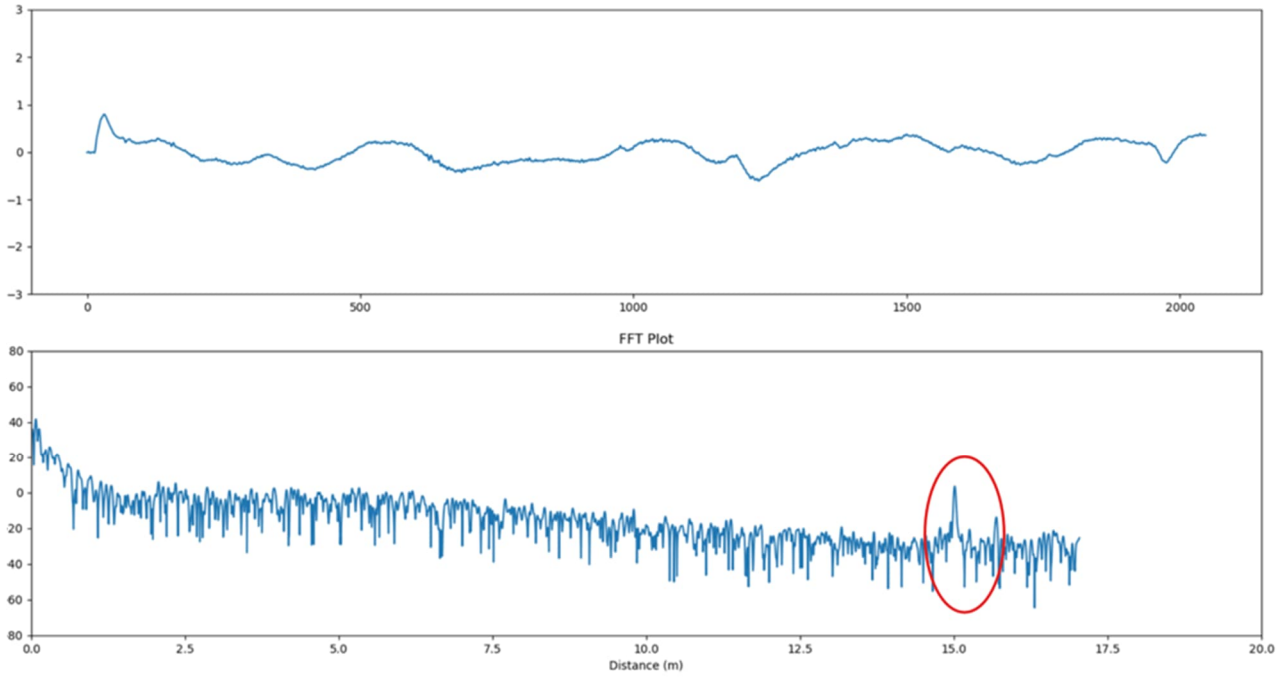


Figure 89 Measurement with corner reflector at 15 m (far-field lens, 50 mm)

In addition, tests were conducted to determine the distance at which it is possible to recognize a person in everyday clothing in a static state. For this purpose, the corner reflector from the previous tests was replaced by a person.

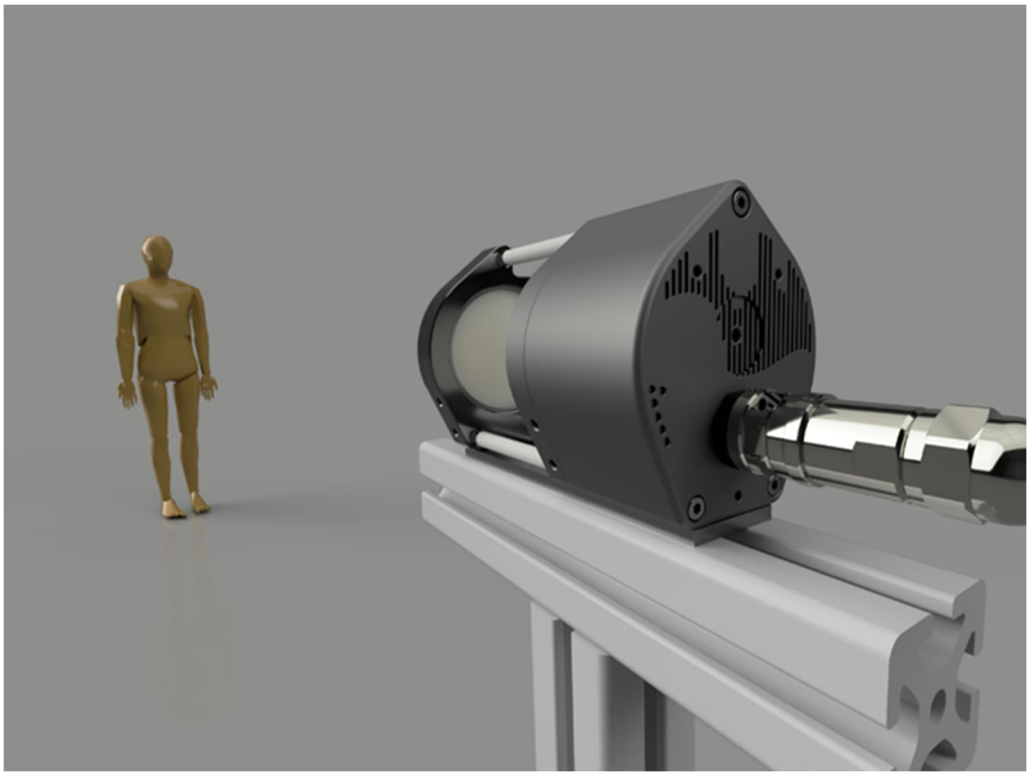


Figure 90 Rendering of test setup for testing the radar+lens configurations capability to detect humans

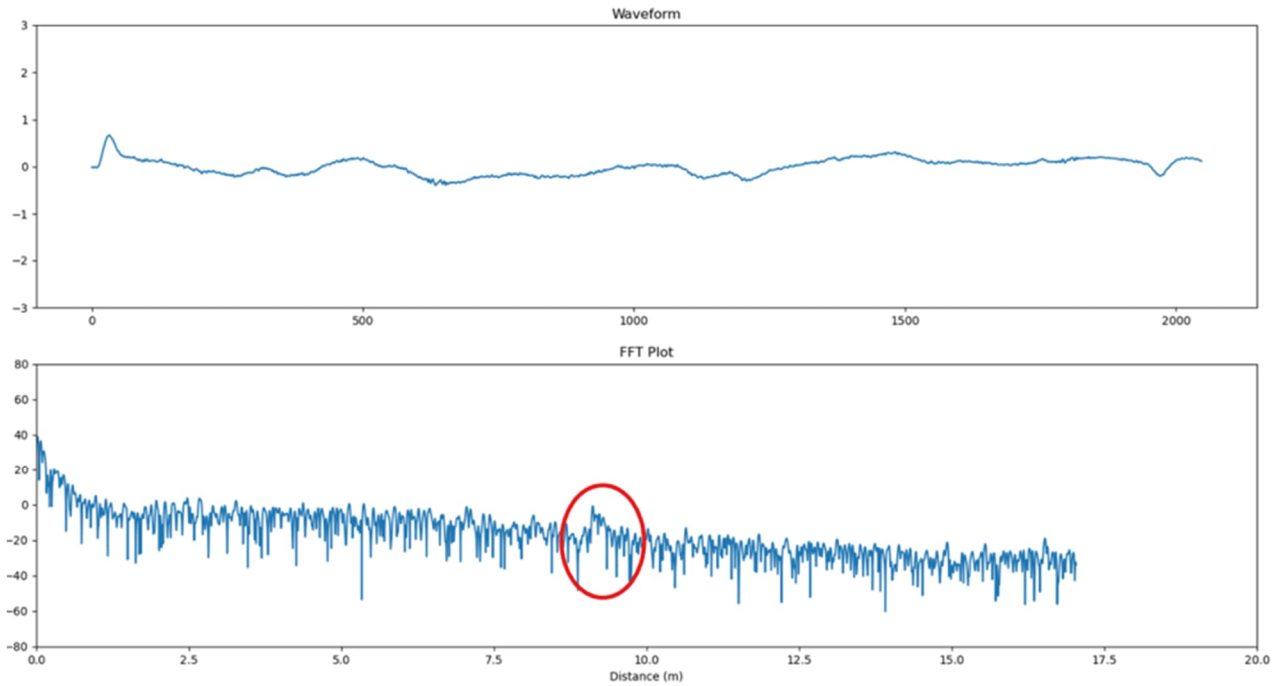


Figure 91 Measurement with a person at 9 m (far-field lens, 50 mm)

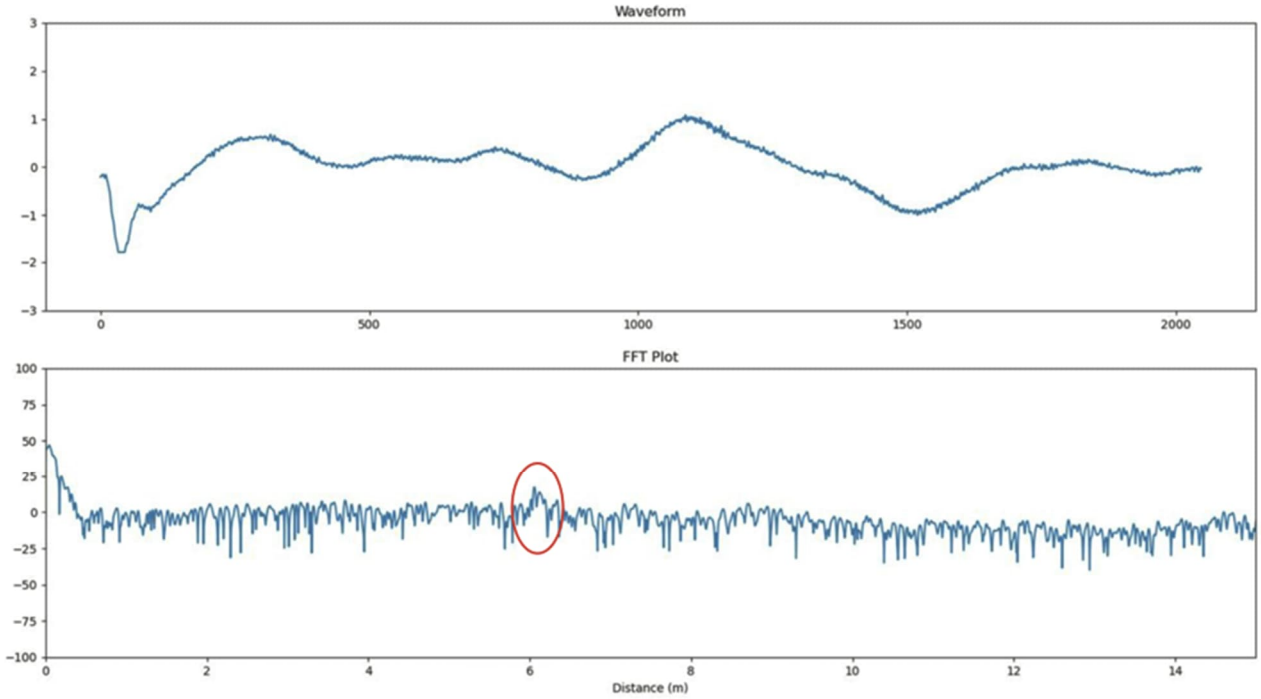


Figure 92 Human at 6 m (hyper-hemispheric lens, 20 mm)

As can be seen from Figure 91 and Figure 92, it is entirely possible to detect a person over a distance of several meters using radar. The elliptical lens achieved the highest range of just under 12 meters in the tests.

However, the problem is the very small aperture angle of the radar beam, which only allows monitoring of a very small area. Here, the trade-off between maximum range and area coverage must be well balanced if dozens of radars are not to be implemented in the final application.
To enable a more accessible method for evaluating different lens configurations, TRILITEC has developed a simple peak detection tool with a graphical user interface (Figure 93) that displays the

distance (calculated on board) to the nearest obstacle. This allows the area covered by the radar to be determined in real time with moving targets.

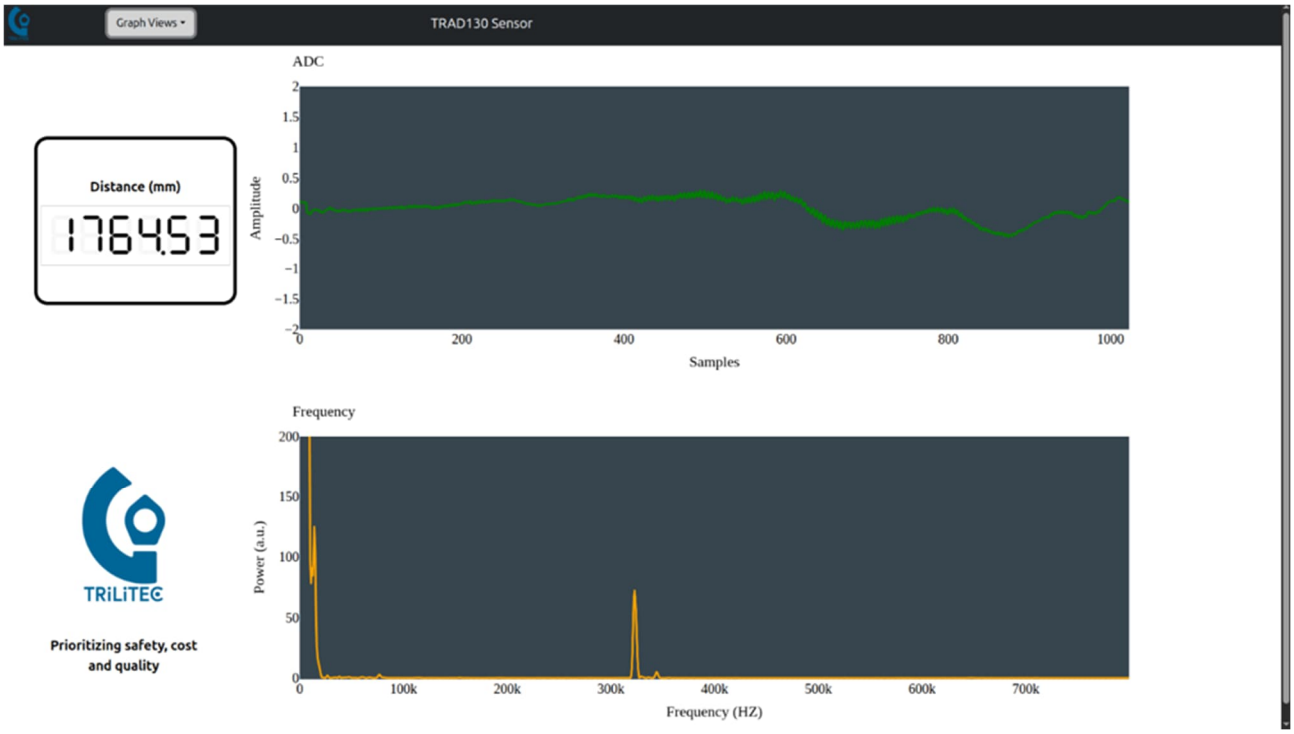


Figure 93 GUI of nearest target estimation tool

Technologies collected and tested from WP2

In the evaluation of this part of the pilot P3 the multiple components used are the results of sensor module development activities from WP2. Under WP2 the components are also individually measured, and the performances are evaluated. However, the above shown full system setup allows the validation of the pilot. Later, this technology will be put on the AMR of the PRODRIVE and will be tested under factory conditions. This will be done in 2nd cycle where more elements from WP2 results shall be added.

5.3.3 Pilot technical KPI measurements

5.3.3.1. P3-SON Evaluation cycle 1 in Son

An updated version of the technical KPI's of the pilot in SON is shown in Table 26. In the rest of the paragraph the methodology and evaluation will be discussed.

Table 26: Updated version technical KPI's

Nr	State of the art	KPI	Target
1.2	Imaging radar-based sensing for human motion detection	The evaluation is based on the gained resolution enhancement and the accuracy to detect body part movements	3D and/or polarized radar. Detection of body movements accurate to 90%

2.3	Cooperative fusion	Improving detection and tracking while keeping complexity low. Easy adaptation to specific operating conditions	Time to first reliable detection of road user 100ms lower than for late fusion, and average precision 5% better in difficult conditions (low illumination, occluded areas)
2.4	Continuous learning	Time-consuming and power-hungry offline learning can be avoided when the machine learning network continuously learns from the input data on the fly	A robust multisensory system, including at least one radar, will be realized without any pre-training, but instead only relying on the continual adaptation of the neural network weights as the system explores the environment
3.3	Distributed learning	I Accuracy of the machine learning component improvement using unsupervised federated learning	15%
4.4	Activity detection	Develop possibility for AGV and robot cells to detect moving humans	Be able to accurately distinguish and avoid moving humans and act upon information on human activity

5.3.3.1.1. Cooperative fusion

In the first cycle, imec conducted experiments mainly on automotive dataset and achieved latency improved by 100ms compared to late fusion. Recording and annotation of multi-sensor sequences in challenging conditions is in progress, while the first evaluations of average precision will be performed in the first months of the second cycle.

5.3.3.1.2. Continuous learning

The current sensor fusion and object recognition based on continuous learning is currently being developed by Imec and Holst. Its evaluation will be performed in the second project cycle.

5.3.3.1.3. Distributed learning

In the first project cycle, imec developed a distributed learning algorithm for object detection. The proposed distributed learning algorithm was tested by deploying 4 object recognition models on different nodes of the intersection and aggregating final decisions from distributed models. Initial experiments showed improvements of 4.2% over the centralized model.

5.3.3.1.4. Activity detection

This KPI relates to the first scenario “Person Recognition”. For the first cycle it is decided to perform this in a controlled lab environment. Whereby the mobile robot just picked up some storage bins and wanted to drive backwards. There are 3 situations that are tested, these are:

1. A person moves in and out of the safety zone
2. Multiple people move in and out of the safety zone
3. Mobile robot and human move in and out of the safety zone.

The results are shown in the figures below:

Figure 94 Test Environment (2 AMR's + (Un)Loading station)

Figure 95 A person moves in and out of the safety zone

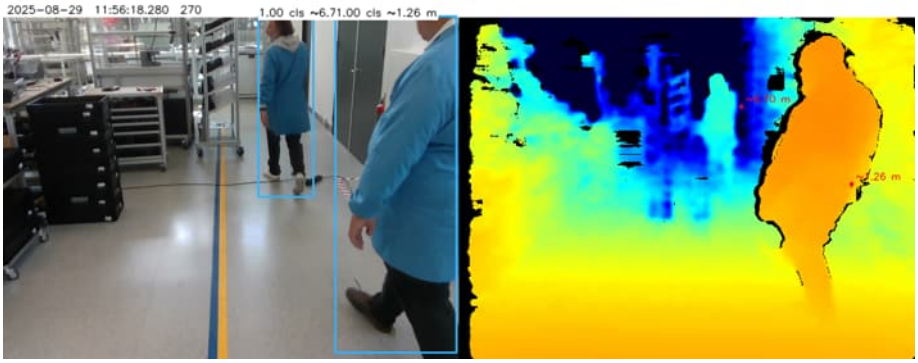


Figure 96 Multiple people move in and out of the safety zone

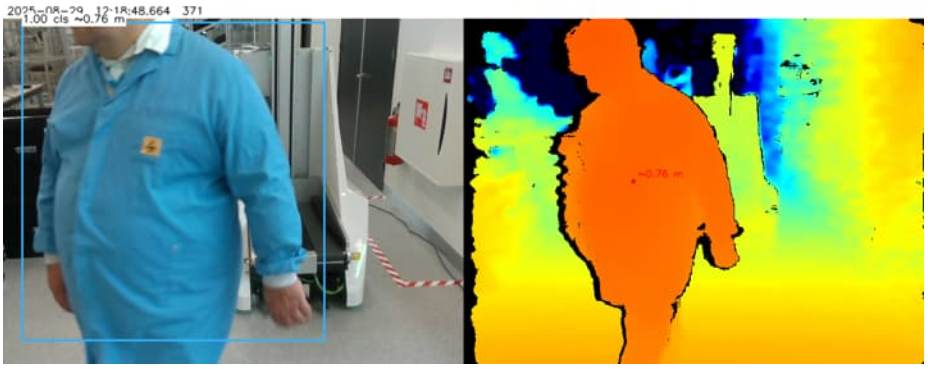


Figure 97 Mobile robot and human move in and out of the safety zone

5.3.3.2. P3-SON KPI Evaluation in Bremen

In the first cycle as mentioned the pilot evaluation was divided into two locations. However, the methodology of pilot evaluation remained the same. The laboratory validations were carried out in Radar Measurement facility of University of Bremen. The cluster of UBREMEN and TRILITEC, together worked on evaluating the developments from WP2 with the following technical KPI are targeted:

Table 27 Technical KPIs for P3-SON

Nr	State of the art	KPI	Planned Target from D5.1	Evaluation Report on the achieved target
1.1	Radar-based sensing	The additional functionalities of the radar sensor will be evaluated at first under laboratory environment for distance measurements for example. Then later in 2nd cycle the system will be tested in real factory environment.	It will be validated that the functionality of the radar sensor is unaltered even when the antenna is moved away from the radar sensor to a different direction/position.	With corner reflectors as Targets the radar unit + software from TRILITEC were tested under two situations i) with a Horn Antenna placed with the radar and ii) with a launcher + dielectric waveguide + lens antenna setup. The 2nd setup allowed the flexible relocating of the radar antenna. It was tested that the radar unit was still able to detect the target and estimate the distance. Further reports and validation graphs are illustrated below.

The pilot evaluation was done by comparing the results between the distance measurement results of the two setups A and B shown in the previous section. The corner reflector is placed at a distance away from the radar. In the graphs shown below it can be seen that the target is detected at a distance of approximately 48 cm from the radar. This is from setup A where the radar is connected to the horn antenna.

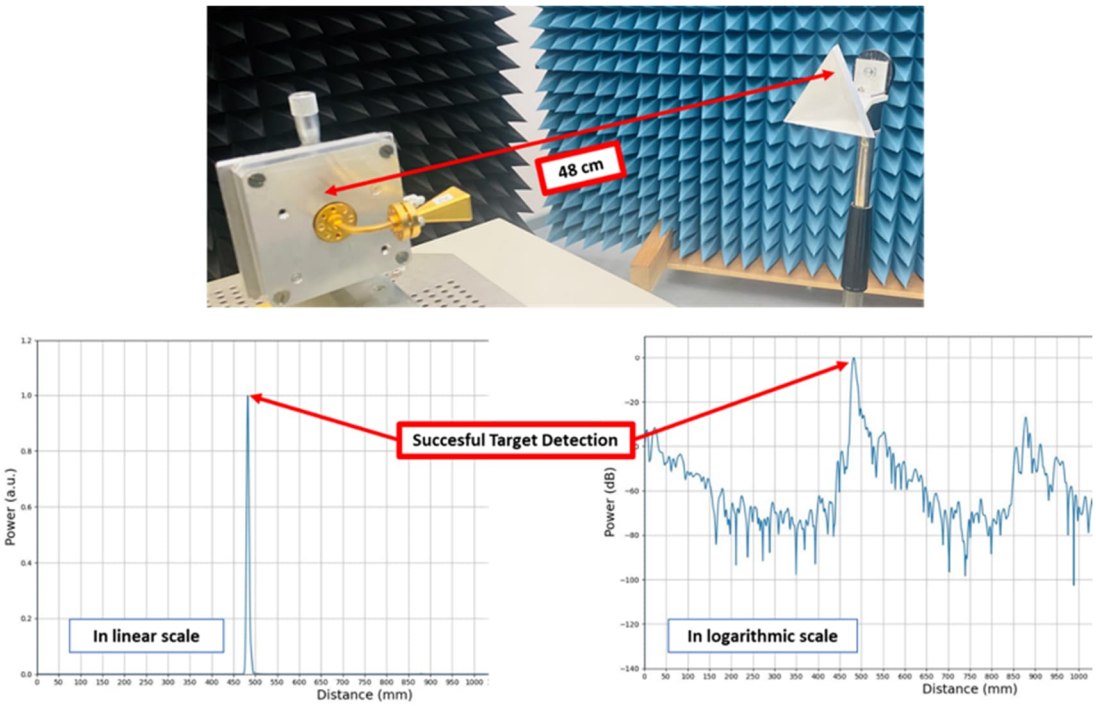


Figure 98 P3-SON Setup A depiction

To compare this with the setup B where the antenna is relocated to a different distance with the help of a dielectric waveguide extension. This is shown below. It can be clearly seen that the detection of the target is still successful even with the extension. There is a shift in the measured distance and that is mainly because the distance measurement in a radar is based on the time delay calculations and in the setup B due to the presence of a dielectric waveguide the waves become slower in this medium. Thereby the measured distance is increased. However, this can be analytically calculated from the length of the dielectric cable and then the radar system can be accordingly calibrated. This is part of the cycle 2 developments and shall be incorporated in the final demo.

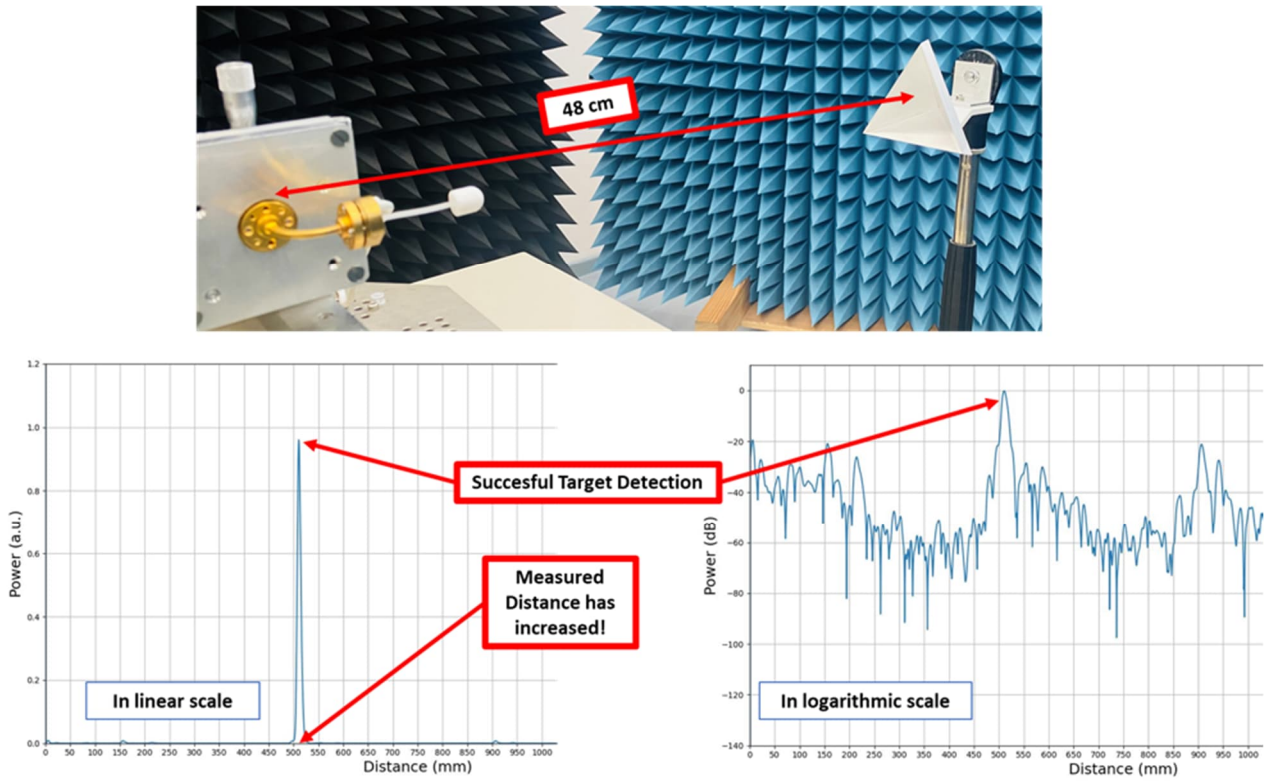


Figure 99 P3-SON Setup B depiction

To fully test this additional functionality of the radar a third setup is made. The picture below shows this setup and its results where the extended cable + antenna is made to peek into a different direction (90 degree rotated).

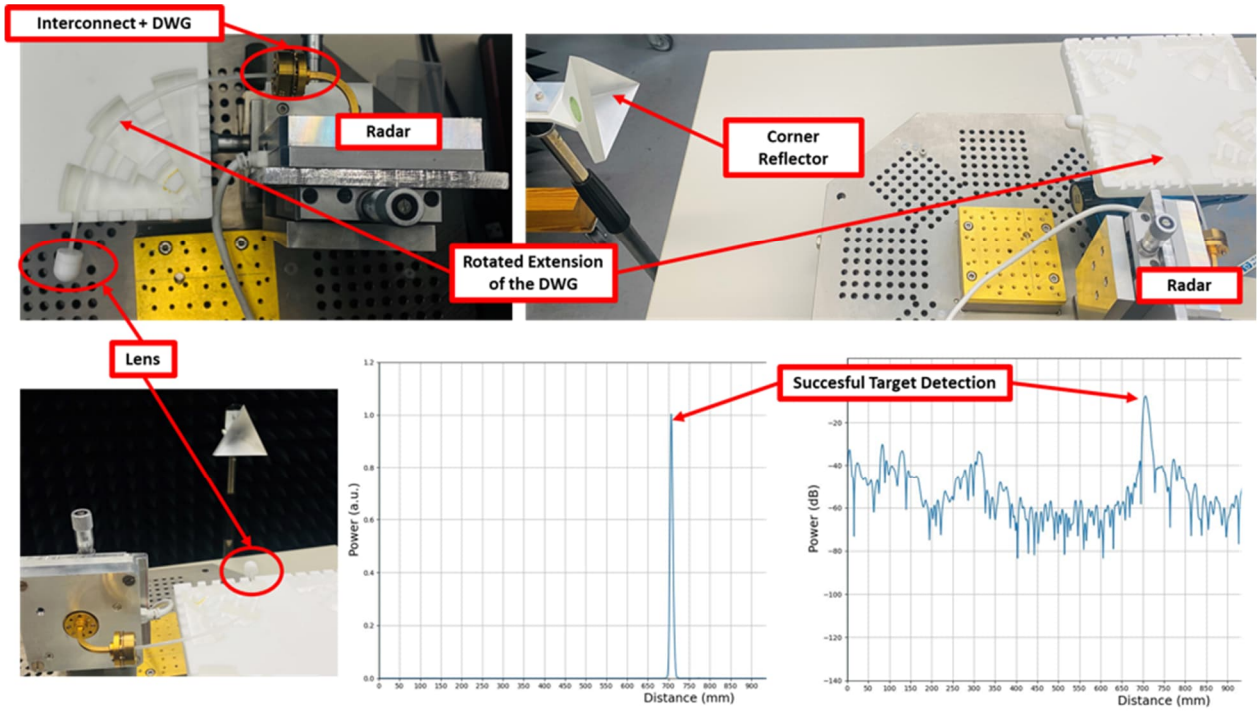


Figure 100 P3-SON Setup C (at 90 degree angle) depiction

In this case there are some other peaks visible before the target, especially in the logarithmic plot which arises from the reflections coming from the support structures or the bending of the dielectric

waveguide. However, their power levels are much lower than the main target peak so it can be easily distinguished. The curves are taken from the software developed by TRILITEC.

As a conclusion to the validation activities, it can be said that the concept is proved that an antenna of a radar unit can also be relocated more flexibly into a different location or even can be pointed to directions that is not on the bore sight of the radar itself. Many such different situations were tested and different waveguide sections, different antennas were tried. The radar unit seemed to be performing quite well with this enhanced feature. It is, however, important to note that due to the presence of the additional dielectric cable and the antenna, slight shifting, broadening or power level reduction were observed that totally co-relate to the theoretical aspects. This is being currently dealt with and soon the calibration routine will be added, which will minimize the effects of this cable extension.

5.3.4 User aspects: stakeholder engagement in pilot development

5.3.4.1. P3-SON Evaluation cycle 1 in Son

The HMI Mobile Robot scenario has been prolonged for the second use case cycle because of small delays in development. It is expected that this part will be tested in the second use case cycle. This is why the stakeholder engagement and user aspects have been prolonged as well. However, in the second use case cycle Prodrive aims to test these KPI's as well in close collaboration with IMEC and Holst. For now, the technical development team was the one in the test environment, but in the second use case cycle it will be operators and factory workers in a real-world industrial environment.

In the lab environment of University of Bremen there are not many stakeholders. The stakeholder description will be similar to the one of Son, because the technologies will be integrated in the second use case cycle.

5.3.4. User-based KPI assessment

The user-based KPI's were related to the "mental stress" of the operator and factory works related to the collaboration with mobile robots. This is similar as demo 3.1. The decision was made to focus first on the technical proof of concept in a controlled lab environment. In this case, the scenario is being played in a specific safe setup which is not representative of the real world. It is not possible to scientifically measure stress levels. Because of that the user-based KPI's will move to the second cycle where it will be reevaluated.

5.3.5. User aspects: Gender/age issues and ethical concerns in P3-SON

P3-BEST, P3-GRA, and P3-SON jointly carried out the ethical reflection, as detailed in section 5.1.6 "User aspects: Gender/age issues and ethical concerns in P3-BEST". We refer to this section for all ethical concerns and user aspects.

5.3.6. P3-SON transition into cycle 2: takeaways and feedback

Highlights of Cycle 1

- Successful implementation of person recognition: mobile robots adapted their movement when humans entered their workspace, ensuring safety without compromising efficiency.
- Sensor fusion and machine learning enabled accurate human detection using radar, LIDAR, RGB, thermal, and DVS inputs.

- Distributed learning algorithms improved object detection accuracy by 4.2% over centralized models.
- Latency improvements of 100ms were achieved in sensor fusion experiments.
- Radar antenna relocation was successfully validated: the TRILITEC radar unit, tested with both a horn antenna and a launcher + dielectric waveguide + lens setup, maintained its ability to detect targets and estimate distances even when the antenna was repositioned.

Lowlights of Cycle 1

- Development delays postponed testing of the HMI Mobile Robot scenario to Cycle 2.
- Technology split between Radar and other components created fragmentation in the architecture, impacting the results.

Intended Changes for Cycle 2 Specification

- Expansion from lab-based testing to real-world industrial applications, incorporating user aspects.
- Centralized pilot to include radar technology and full sensor box deployment on AGVs.
- Continued development and evaluation of continuous learning for object recognition.
- Integration of WP2 results into factory conditions at Prodrive, validating the full system setup.

6. Summary and next steps

6.3. Conclusions and takeaways from cycle 1 evaluation

The purpose of the validation undertaken in cycle 1 was to evaluate whether the demonstrators met the expectations laid down in the Use Cases and how well they performed with respect to the requirements set in the project. 16 pilots have been presented according to the general three Use Cases (Health monitoring, Driving monitoring and Safe Interaction with robots) that focus on different aspects to be solved with distributed sensing and intelligence solutions developed in the work packages WP2 through WP5 in DistriMuSe. Despite some initial difficulties in some of the pilots (e.g., most of the work in the pilots in Finland was delayed by problems in the national funding that were solved too late to adhere to the initial calendar) all pilots showed remarkable results and good validation of the expected progress during the first period of work. Additionally, all pilots provided an overview of the lessons learnt and an outlook to the 2nd project cycle project.

6.4. Transition into cycle 2: overall takeaways

DistriMuSe cycle 1 has been focused on constructing useful proof of concept technologies and validating their usefulness in environments similar to those in which these would be applied in the real world. Emphasis, then, has been predominantly technical: selection of useful technologies, integration into practical demonstrators and very limited tests to validate their potential usefulness in real world scenarios.

In addition, a number of limitations have appeared that have minimised our full potential in this phase:

- Delays in the initial set-up in all pilots involving Finnish partners due to national funding coming later than planned. This has disrupted pilots around the project but principally felt more strongly in pilots led by these Finnish partners (such as P1-KUO, P1-KSL and P2-TMP). With the funding arriving very late into the planning of cycle 1, there was limited activities that could be undertaken to flesh out these pilots into their full realization. So, it was decided that the work during the evaluation phase of cycle 1 would be mostly used for detailed planning and that real evaluation would be delayed and largely conducted during cycle 2. This has implications on both WP5 and WP1 which will be further explained in future deliverables of both work packages, such as the upcoming D1.4.
- Some limitation on the access to important equipment and infrastructure limited pilots in use cases 1 and 2 such as P1-LL and P2-BRU. In these pilots, work has been performed but not all of the results have been achieved. This will be corrected during cycle 2. In addition to this initial work, useful data has been captured to help in the continuing developments.

Other than these limited deviations, work in WP5 during the first cycle concludes with this document. In the cycle 2 we will deepen into the more technical approaches (e.g., using more complex algorithms and more focused training data using the cycle 1 results) as well as emphasize the user-oriented aspects of the pilots, which have only been explored superficially in the first cycle. The second cycle of DistriMuSe is expected to produce the promised advances over the current State of the Art as well as produce useful insights on the user aspects associated to the pilots, such as technology acceptance, social impact and trust. It is expected that, beyond the pure application of technology, this emphasis will round up the results so that adoption can be done in the exploitation phase with full confidence not only the usefulness of technology but also its full impact on humans.

6 References

Bruser, C. ; Kortelainen, J.M. ; Winter, S. ; Tenhunen, M. ; Parkka, J. ; Leonhardt, S.: Improvement of Force-Sensor-Based Heart Rate Estimation Using Multichannel Data Fusion, Biomedical and Health Informatics, IEEE Journal of Volume: 19, Issue: 1, DOI: 10.1109/JBHI.2014.2311582, Publication Year: 2015, Page(s): 227 – 235

Peng, RC., Li, Y. & Yan, WR. A correlation study of beat-to-beat R-R intervals and pulse arrival time under natural state and cold stimulation. Sci Rep 11, 11215 (2021). https://doi.org/10.1038/s41598-021-90056-2

Presta, R., Tancredi, C., De Simone, F., Girau, R., Monteleone, A., & Viero, F. (2025, June). Augmenting Driver Situation Awareness through Distributed Sensing and Driver Adaptive Interfaces: a Research Framework. In International Conference on Human-Computer Interaction 2025. To appear.

Zambudio Martínez, M., Marin-Perez, R., & Skarmeta Gomez, A. F. (2025). Development of an In-Vehicle Intrusion Detection Model Integrating Federated Learning and LSTM Networks. Information, 16(4), 292. https://doi.org/10.3390/info16040292