SCIENCE CHINA
Information Sciences

• NEWS & VIEWS •

June 2026, Vol. 69, Iss. 6, 167301:1–167301:2
https://doi.org/10.1007/s11432-026-4923-1

PML-CellularEye: a multi-modal experimental dataset for 6G ISAC research

Ziguo ZHONG[1], Yongming HUANG[1,2*], Huazhou HOU[1], Fanfei XU[1,2], Haisheng FENG[1], Shengheng LIU[1,2] & Xiaohu YOU[1,2*]

[1]*Purple Mountain Laboratories, Nanjing 211111, China*
[2]*School of Information Science and Engineering, Southeast University, Nanjing 210096, China*

Received 25 March 2026/Revised 14 April 2026/Accepted 16 April 2026/Published online 12 May 2026

Citation Zhong Z Z, Huang Y M, Hou H Z, et al. PML-CellularEye: a multi-modal experimental dataset for 6G ISAC research. Sci China Inf Sci, 2026, 69(6): 167301, https://doi.org/10.1007/s11432-026-4923-1

On November 26, 2025, Purple Mountain Laboratories (PML) released the CellularEye v1.0.0 dataset that offers the first batch of large-scale multi-modal data acquired from its newly built experimental testbed for 6G radio access network (RAN) and integrated sensing and communication (ISAC) research.

It is widely regarded that the evolution toward the sixth generation (6G) mobile communication represents a fundamental architectural shift from connectivity enhancements to the creation of intelligent systems capable of simultaneous communication, environmental sensing, and edge computing [1]. ISAC emerges as a transformative capability, which turns the communication network into a distributed sensor array, and its synergy with AI-native radio access networks is poised to redefine 6G's core [2]. Under this vision, data become pivotal and many efforts have been devoted to creating synthetic [3–5] or empirical [6] datasets for communication study, yet there is still a lack of large-scale and multi-modal experimental datasets dedicated to ISAC research.

To meet this goal, PML has been developing RAN testbeds with non-intrusive, multi-modal and scalable data acquisition capabilities. The center site for the ISAC testbed is shown in Figure 1(a), a field occupying a paved area of 228 m × 70 m. Currently, this site is equipped with twelve mmWave active antenna units (AAUs), four sub-6GHz remote radio units (RRUs), two standard baseband processing units, four ISAC data processing and edge computing units, an optical switching network tailored for accurate timing (required by 5G/6G RAN standards), and a 100 TB data storage facility. The experimental RAN is deployed with a cell-free architecture [7] to enable multi-station distributed collaboration. For the study on environment sensing, the testbed also comprises two on-premises drone airports, two optoelectronic systems with both visible-light and infrared vision, and several types of vehicles and robots for testing purposes. In addition, a micro meteorological station is installed for on-site weather monitoring, and multiple global navigation satellite system (GNSS) receivers are deployed to provide in situ synchronization and positioning services.

Designed to support data-centric research such as ISAC, the testbed features several key capabilities.

• Non-intrusive data acquisition: The testbed is capable of collecting air-interface raw data (i.e., I/Q data) without interrupting the normal operation of gNBs. The fidelity of the data is assured through extensive validation using bit-exact matching methods. Built on a modular architecture, the data acquisition system can scale easily with an increasing number of RAN nodes[1)].

• Accurate time synchronization: By designing optimized clock distribution networks, the testbed ensures precise time synchronization among gNB functional units (at the level of 10s of ns and 10 ppb) to comply with 3GPP standards. For non-communication devices, their operations and outputs are also synchronized to ensure temporal and spatial consistency of the collected data.

• Automated data management: An automated data management framework has been established, featuring procedures for scheduled drone flights, coordinated multi-modal data acquisition, and on-site data cleansing, formatting, archiving, and status reporting. The processed data not only provide rich information for various ISAC use cases [8] but also facilitate cross-modality verification, dataset labeling, and interpretation.

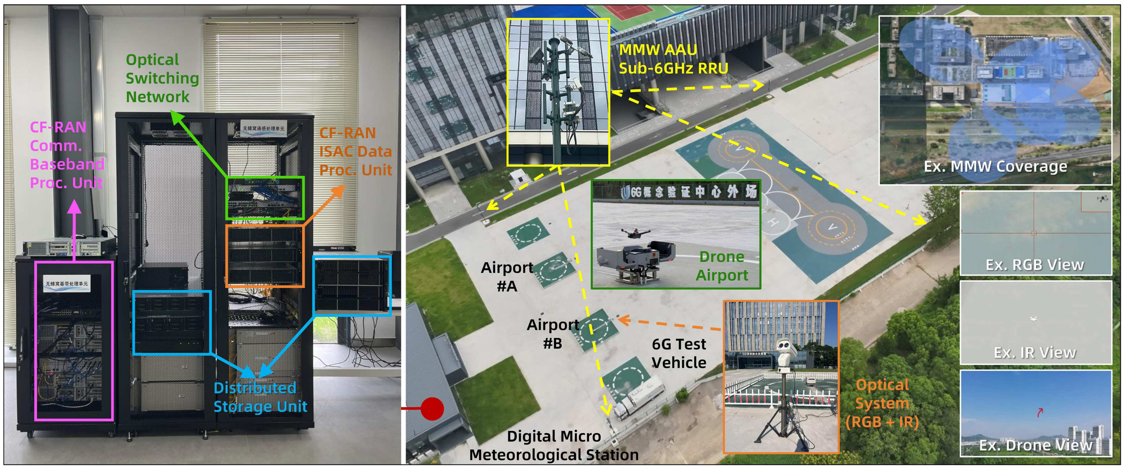
Further technical details about the testbed, including hardware configurations, signal and data specifications, benchmark methods, and so forth, are available on the dataset website at http://pmldatanet.com.cn/dataapp/multimodal.

The first batch of data, CellularEye v1.0.0 as shown in Figure 1(b), has been released for open access. The CellularEye dataset is designed for environment perception using radio access networks, and aims to bridge the gap between ISAC theories, algorithms, and the physical world. Key features of the dataset are summarized below.

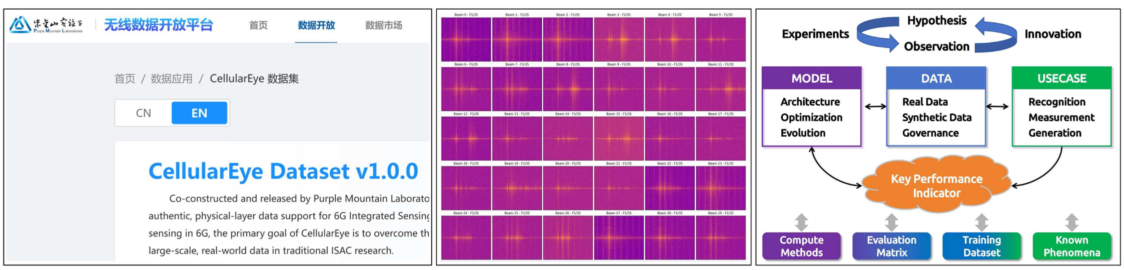
• Scenarios and operations: The CellularEye dataset supports research on ISAC use cases, including "low-altitude" drone detection and environment sensing. Data collection was organized using a RAN with multiple transmit and receive points (mTRP) comprising four mmWave sites, alongside simultaneous optical video streams and real-time weather data. Each round of data acquisition lasted 10 s and was carried out four times daily (at 00:00, 06:00, 12:00, and 18:00). During specific collection tasks, one or

* Corresponding author (email: huangym@seu.edu.cn, xhyu@seu.edu.cn)

1) Commercial off-the-shelf (COTS) 5G/5G-A gNB devices are utilized for research purpose.



(a) PML's ISAC Testbed: mmWave AAUs, sub6-GHz RRUs, baseband & data proc. units, opticals, drones & weather station



(b) CellularEye Dataset v1.0.0: access website (left), mmwave data visualization (middle), data-centric supports (right)

Figure 1 (Color online) An overview of PML's testbed and CellularEye dataset for 6G ISAC research.

two DJI Matrice 3D drones performed coordinated flights to emulate aerial intrusion events; their onboard real-time kinematic (RTK) receivers provided ground-truth coordinates. More details regarding the scene and dataset can be found on the website.

• Data organization and long-term evolution: Data records are archived by collection sequence (i.e., the start time of the task). Specifications regarding system configurations and data formats are detailed in the accompanying document. Script tools are provided to help users read, visualize, and understand the data, as shown in the middle part of Figure 1(b) with an example. Currently, the CellularEye dataset contains approximately 3 TB of multi-modal experimental data. As a long-term goal, a 100 TB dataset (physically collected real data) is expected to be established within the next two years to cover more use cases and RAN settings.

• Usage and community support: The dataset has been employed by research teams from multiple organizations for 6G study. The real, multi-modal data can be used for both classical ISAC algorithm development and data-driven neural network modeling for broader applications (e.g., AI-RAN [2]). As illustrated in Figure 1(b), a suite of data-centric supporting resources will also be provided to the community, including reference models, use-case algorithms, compute libraries, and data management tools. Furthermore, this platform will serve as a testbed to explore and evaluate innovations such as new waveforms [9], synthetic data enrichment [10], and AI edge technologies [11].

Conclusion. This article reports the CellularEye dataset, a large-scale, multi-modal experimental dataset designed to support ISAC research. This open resource provides real-world data that addresses the scarcity issue currently limiting sensing and perception studies. As a next step, the testbed and dataset will be expanded to include more network configurations, additional hardware (such as reconfigurable intelligent surfaces) and joint research projects. We hope that these efforts will facilitate broader 6G research in the community.

Access methods. The CellularEye dataset can be accessed from http://pmldatanet.com.cn/dataapp/multimodal. To comply with mandatory regulations, users who wish to download the data are welcome to contact the authors for account information.

Acknowledgements This work was supported in part by Mobile Information Networks-National Science and Technology Major Project (Grant Nos. 2025ZD1302100, 2026ZD1306100).

References

1 You X H, Wang C-X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. Sci China Inf Sci, 2021, 64: 110301
2 You X H, Huang Y M, Zhang C, et al. When AI meets sustainable 6G. Sci China Inf Sci, 2025, 68: 110301
3 Yu L, Zhang J H, Han S F, et al. BUPTCMCC-6G-DataAI+: a generative channel dataset for 6G AI air-interface research. Sci China Inf Sci, 2025, 68: 197301
4 Cheng X, Huang Z, Yu Y, et al. SynthSoM: a synthetic intelligent multi-modal sensing-communication dataset for Synesthesia of Machines (SoM). Sci Data, 2025, 12: 819
5 Chang S, He J S, Huang S, et al. CSRD2025: a large-scale synthetic radio dataset for spectrum sensing in wireless communications. 2025. ArXiv:2508.19552v1
6 China Academy of Information and Communications Technology (CAICT). Mobile Communication Open Dataset. https://www.mobileai-dataset.cn
7 Wang D M, You X H, Huang Y M, et al. Full-spectrum cell-free RAN for 6G systems: system design and experimental results. Sci China Inf Sci, 2023, 66: 130305
8 Zhang J F, Lu W D, Xing C W, et al. Intelligent integrated sensing and communication: a survey. Sci China Inf Sci, 2025, 68: 131301
9 Xiao L, Li S, Qian Y, et al. An overview of OTFS for Internet of Things: concepts, benefits, and challenges. IEEE Internet Things J, 2022, 9: 7596–7618
10 Chi G X, Yang Z, Wu C S, et al. RF-diffusion: radio signal generation via time-frequency diffusion. In: Proceedings of the 30th Annual International Conference on Mobile Computing and Networking, 2024. 77–92
11 Yang T T, Huang N, Sun L, et al. AI Edge: mobile information service infrastructure for intelligent applications (in Chinese). Sci Sin Inform, 2025, 55: 2637