Development and Sharing of a Multi-Modal Indoor UAV Dataset for PHM Research

Chun Fui Liew, Gengyu Li, Akira Osaka, Samir Khan, Naoya Takeishi and Takehisa Yairi

Research Center for Advanced Science and Technology (RCAST), The University of Tokyo
{*liew, yairi, ntake, kviz3671, li-gengyu, akr-osaka*}*@g.ecc.u-tokyo.ac.jp*

ABSTRACT

With the rapid expansion of unmanned aerial vehicle (UAV) applications, ensuring reliable and safe operation has become a pressing challenge. In this paper, we present a modular quadrotor-based data acquisition platform designed to capture rich, high-fidelity data under motion capture guidance. Our system integrates conventional flight telemetry with detailed vibration and temperature measurements, user input logs, and precise 6D motion tracking. This offers an comprehensive view into the UAV's physical and control state. We describe our systematic process for data cleaning, organization, and exploratory analysis, laying the groundwork for robust prognostics and health management (PHM) research. To illustrate the platform's potential, we implement supervised and semi-supervised models for anomaly detection and fault identification. We release the dataset, synchronized flight videos, and analysis code to accelerate UAV health-monitoring research and collaboration.

1. INTRODUCTION

Unmanned aerial vehicles (UAVs) are increasingly employed in diverse indoor applications, spanning from warehouse logistics to inspection tasks, where reliable operation is paramount to mission success. Nevertheless, the dynamic nature of quadrotor flight, coupled with intricate mechanical and electrical subsystems, presents substantial challenges to their long-term reliability and safety. Prognostics and Health Management (PHM) presents a promising pathway to address these challenges through condition-based monitoring, anomaly detection, and remaining useful life estimation. However, the development and validation of PHM methods critically depend on high-quality, representative datasets, which remain scarce in the context of UAVs, particularly for indoor scenarios where precise ground truth is available.

To address this need, we introduce a novel indoor UAV data acquisition platform centered around a motion capture system that enables sub-millimeter precision in trajectory tracking. Our platform captures rich, synchronized, multi-modal sensor data—including gyroscope and accelerometer readings, motor RPM, ESC and motor temperatures (via both contact-based and infrared sensors), vibration signals, battery voltage and current, and RC transmitter inputs. Additionally, we collect 6D motion capture data comprising 3D position and 3D velocity. A key feature of our design is the rigidly mounted flight controller board, which allows direct acquisition of vibration data from the airframe itself, providing unfiltered insights into mechanical integrity. By combining precise motion ground truth with diverse onboard telemetry, our system provides a rare, high-resolution view into the health of UAV subsystems during real-world operation. We detail the data collection and cleaning procedures, and present initial exploratory analyses to ensure dataset quality. Furthermore, we implement baseline supervised and self-supervised AI models for anomaly detection and health estimation. To foster open research, we release the complete dataset along with synchronized flight videos and sample data processing scripts in Python.

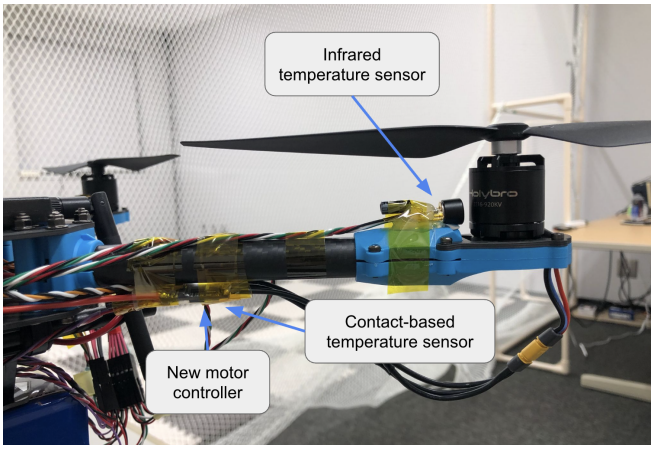

Figure 1. Quadrotor arm: an IR sensor targeting the motor bell and a contact sensor on the ESC provide complementary surface and component temperature measurements for PHM.

2. RELATED WORKS

Data-driven PHM has a long history spanning aerospace and industrial systems, with broad surveys consolidating classical paradigms such as distance-based, density-based, and model-based approaches, along with their time-series variants (Chandola et al., 2009). Widely used algorithms include One-Class SVM for novelty detection (Schölkopf et al., 2001), Local Outlier Factor for density-based outliers (Breunig et al., 2000), and Isolation Forest for scalable, subsampled anomaly scoring (Liu et al., 2008). In prognostics, the NASA C-MAPSS turbofan benchmark popularized run-to-failure sensing for health estimation and RUL modeling and remains a de facto yardstick for method comparisons (Saxena and Goebel, 2008).

Within spacecraft operations, Yairi *et al.* proposed a probabilistic clustering and dimensionality-reduction framework for satellite housekeeping telemetry, demonstrating effective health monitoring without explicit fault labels—an approach that has influenced subsequent unsupervised PHM pipelines (Yairi et al., 2017). For UAVs, much of the public data targets state estimation and perception rather than health monitoring; e.g., the EuRoC MAV dataset has become a standard for visual–inertial odometry (Burri et al., 2016), with VINS-Mono a commonly used baseline (Qin et al., 2018). In contrast, open datasets that couple rigid-mount vibration, dual-mode temperature sensing (contact and IR) of propulsion components, and motion-capture ground truth are scarce—modalities directly aligned with PHM objectives. Our work contributes an indoor UAV dataset tailored for PHM, alongside baseline supervised and self-supervised detectors, evaluated under a flight-grouped protocol that strictly separates flights between training and testing to ensure fair and realistic benchmarking.

3. OUR PLATFORM

We developed a custom quadrotor platform to support comprehensive PHM studies. The system logs system-level telemetry (RC inputs, raw IMU, barometer, magnetometer), battery/power metrics (voltage, current, power state), motor/ESC status, CPU load and system usage, and state estimates (attitude, position, velocity), as well as structural and thermal health signals directly relevant to fault diagnosis.

3.1. Vibration Sensors

To capture true airframe dynamics, one sensor board is rigidly mounted to the frame (no soft-mount). This configuration records three-axis vibrations that reflect structural loads and resonance, enabling early detection of rotor imbalance or frame damage.

3.2. Temperature Sensors

We instrument the propulsion system with both contact and infrared sensing. Four DS18B20 contact sensors are attached to the ESCs (plus a central ambient reference), and five MLX90614 infrared sensors monitor motor surface temperatures (four motors + one central reference), supporting timely detection of thermal overloads that could compromise reliability and safety.

4. DATA COLLECTION

We fuse three sources of flight data: (i) PX4 flight-controller logs, (ii) an auxiliary controller that records structural-health signals (vibration and temperature), and (iii) an external motion-capture (MoCap) system providing high-precision 6-DoF ground truth for indoor flights.

PX4 stream. Beyond control inputs and raw sensors, PX4 supplies battery/power telemetry (voltage, current, power state), propulsion metrics (motor/ESC rotational speed, voltage, temperatures), CPU load/system usage, processed state estimates (attitude, position, velocity), and flight-mode and system-status transitions (manual, autonomous, failsafe). These fields enable post-flight analyses that link pilot intent and vehicle response to health and performance outcomes.

Auxiliary stream. The frame-mounted board contributes three-axis vibration and detailed thermal measurements of ESCs and motors, capturing structural loads and propulsion heating directly at the source.

MoCap stream. Motion-capture trajectories provide a high-fidelity reference for dynamic analysis and validation of onboard estimates.

Synchronization. Each logger runs on an independent clock, so constant offsets and slow drifts can occur. We apply a lightweight, data-driven alignment: for each modality pair we identify the time shift that minimizes the sum of squared differences—equivalently, maximizes correlation—over overlapping signals, after resampling to uniform grids and filling short gaps via interpolation. This method adjusts only the initial offset, rather than performing step-wise re-alignment as in DTW. Although this raises the issue of potential slow drifts, such effects are negligible in our ten-minute experiments; in longer settings they could be mitigated by periodic resynchronization, resampling to a reference clock, or adaptive correction. The resulting common time base enables cross-modal analyses—relating RC commands to responses, linking vibration/temperature to propulsion loading, and benchmarking state estimates against MoCap—while maintaining the timing accuracy necessary for PHM.

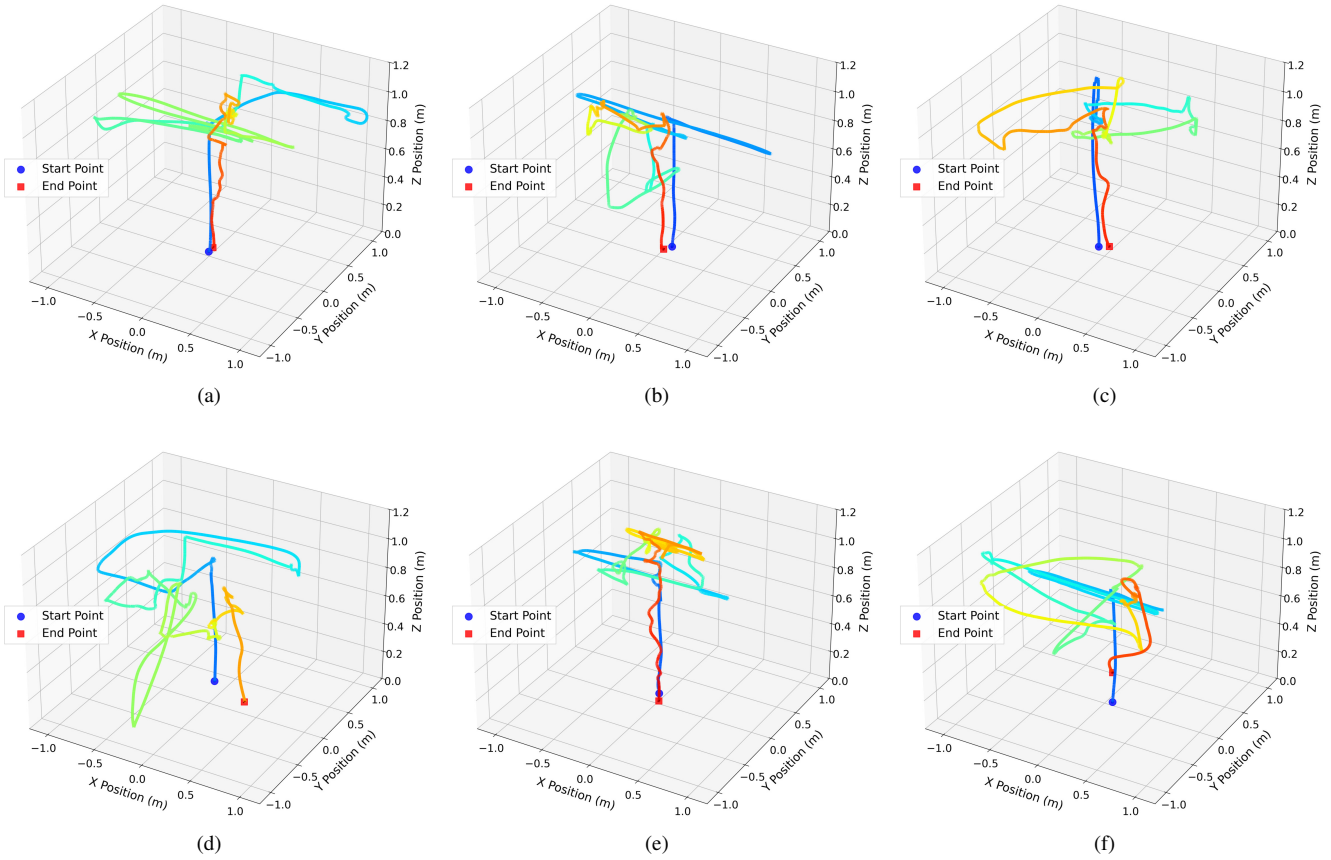

Figure 2. 3D motion-capture trajectories for six flights covering the same indoor workspace. Top row (a–c): **Normal** runs. Bottom row (d–f): **Anomalous** runs — (d) Imbalance, (e) Severe Imbalance, (f) Cracked propeller. Curves are colored by time progression; the blue circle marks the start and the red square marks the end. Axes are in meters (altitude up to $\sim$1.2 m). The comparable spatial envelopes across conditions indicate that faults were induced under similar flight volumes and pilot behaviors, allowing subsequent analysis to attribute differences to health state rather than trajectory coverage.

5. DATA CLEANING PROCESS

To ensure reliable downstream analysis on our multi-modal logs, we apply a concise, uniform preprocessing pipeline. Each stream is first placed on a common time base via data-driven alignment (time–shift search that maximizes cross-correlation on overlapping signals). After alignment, we **subsample by a fixed factor of 1/100**, yielding a common rate of $\approx$200 Hz across modalities, which preserves flight dynamics while reducing volume. Short gaps introduced by asynchronous logging are then imputed with **forward-fill** and **backward-fill** (`ffill`/`bfill`); we keep longer gaps as missing to avoid artifactual data.

Figure 2 visualizes representative 3D trajectories for six datasets (top: normal; bottom: abnormal), all flown by a pilot with randomized maneuvers. Table 1 summarizes the six datasets used in this study and the condition labels employed in our experiments.

Table 1. Overview of the six datasets, condition labels, number of flights, and duration of each flight .

Label used in paper	Number of flights	Duration of each flight
Normal (1)	3	About 1 minute
Normal (2)	3	About 1 minute
Normal (3)	3	About 1 minute
Imbalance (propeller)	3	About 1 minute
Severe-Imbalance (propeller)	3	About 1 minute
Crack (propeller)	2	About 1 minute

6. EXPLORATORY EXPERIMENTS

6.1. Representative signals (overview)

Figure 3 overlays a small set of indicative channels across the flight. The vibration axis (e.g., `sensor_accel_fifo_z`) shows a strong amplitude increase during powered flight, tracking `esc_rpm` and battery current. RC input steps appear as discrete changes in `input_rc_values[2]`, which induce attitude and lateral acceleration responses (`attitude[2]`, `vehicle_local_position_0_ay`).

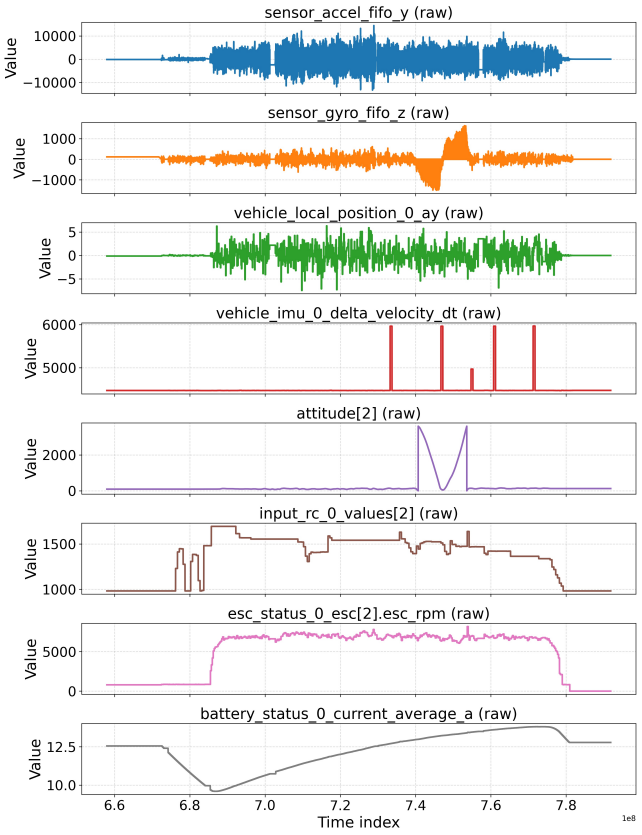

Figure 3. A few representative synchronized channels from one flight (subsampled to 200 Hz).

Near takeoff/landing the current and RPM ramp up/down and the vibration envelope follows accordingly. This view ties pilot commands to propulsion load and structural response, and motivates vibration/temperature features for PHM.

6.2. Hierarchical Clustering

Motivation. Group channels with similar temporal behavior so we can (i) spot redundant/degenerate signals and (ii) pick a small set of representatives for analysis and modeling.

Method. We normalize each channel to have zero mean and unit variance, and then perform agglomerative hierarchical clustering on the full time profiles using Euclidean distance and Ward linkage. This procedure is mathematically equivalent to correlation distance up to a constant scaling, meaning that negatively correlated signals (e.g., correlation = –1) are treated as maximally distant and only positively correlated profiles cluster together. *Pilot-command channels* (e.g., RC inputs, arming/mode flags) are excluded to avoid clustering driven by externally commanded steps rather than intrinsic sensor/propulsion dynamics. The dendrogram is cut into six clusters; each panel plots the normalized traces of all channels in that cluster.

Result (Fig. 4). Cluster 1 (2 signals) shows slowly varying trend-like channels; Cluster 2 (11) contains stepwise/high-variance traces with intermittent spikes (typical of actuator or mode transitions); Cluster 3 (182) aggregates many low-amplitude, quasi-stationary sensors with occasional steps (high redundancy); Cluster 4 (1) is a high-energy vibration channel with a clear flight envelope; Cluster 5 (1) is near-constant (likely a degenerate/flag channel); Cluster 6 (1) is a binary/latched signal (e.g., arming or mode flag). The separation highlights functional groups (vibration, flags, actuators) and exposes redundant or uninformative channels, enabling dimensionality reduction by selecting one representative per cluster.

6.3. Principal Component Analysis

Motivation. We use PCA to reveal dominant modes and temporal structure in the synchronized, multi-modal signals, and to gauge redundancy across channels before building detectors.

Method. After time alignment and 1/100 subsampling (≈200 Hz), each channel is normalized to zero mean and unit variance (z-scored) and then stacked into a time×feature matrix. PCA is fit separately to each dataset; we report the variance spectrum and visualize the time trajectory in the first components (color = normalized time).

Findings (Fig. 5–8). Figures 5 and 6 are shown for normal (nominal) flight data, whereas the projection plots (Fig. 7–8) cover both normal and anomalous cases. (i) *Variance spectrum:* PC1 explains $\approx$ 26% of the variance, PC2 $\approx$ 7%, and PC3 a few percent; the cumulative curve shows an elbow around 5–7 components and reaches $\sim$ 0.75 by 25 PCs (Fig. 5). This indicates strong low-dimensional structure with moderate residual complexity. (ii) *Time–domain PCs:* The panels show the first ten *scores* (PC1–PC10) versus time after projecting the z-scored multivariate signal onto the PCA basis (Fig. 6). Consistent with the variance spectrum (PC1 ≈26%, PC2 ≈7%), **PC1** captures the slow flight envelope: a rise at takeoff, a broad plateau during powered flight, and a decay on landing (a surrogate for overall thrust/propulsion load). **PC2–PC3** capture related but distinct transients around maneuver phases: PC2 emphasizes the broader acceleration–deceleration trend, while PC3 highlights sharper, localized excursions that resemble attitude bias effects. **PC4** isolates a short, high-amplitude burst (step-like actuator/mode transition). Higher components **PC5–PC10** contain higher-frequency, quasi-stationary content with intermittent spikes, consistent with vibration/attitude oscillations modulated by control inputs. (iii) *2D projections:* The PC1–PC2 plots form smooth arcs/loops colored by time (Fig. 7), consistent with flight phases (takeoff, maneuvering, landing). Different flights exhibit similar shapes, suggesting a shared manifold of nominal dynamics. (iv) *3D projections:* Including PC3

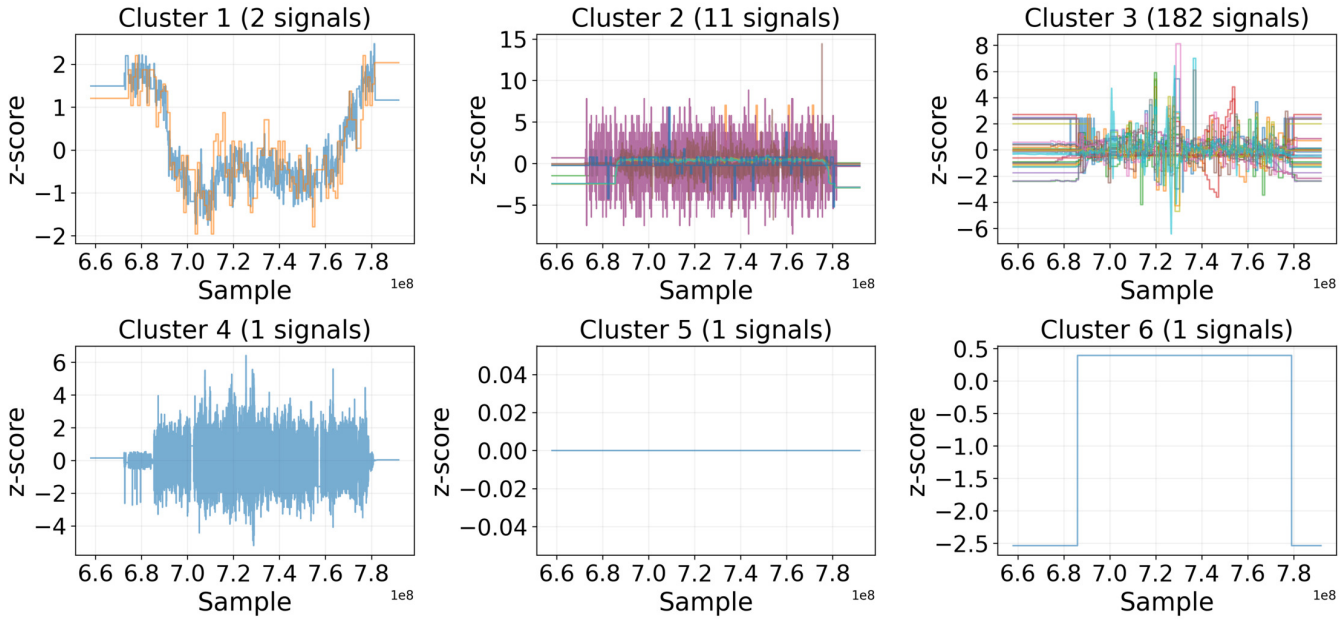

Figure 4. Agglomerative hierarchical clustering of z-scored channels (Euclidean distance, Ward linkage). Each panel shows all signals in a cluster; the title lists the cluster size. Cluster 1: slowly varying trend-like signals; Cluster 2: stepwise/high-variance traces with intermittent spikes; Cluster 3: many low-amplitude, quasi-stationary sensors (high redundancy); Cluster 4: high-energy vibration with a clear flight envelope; Cluster 5: near-constant zero in this dataset—kept intentionally since the same channel is non-constant in other datasets; Cluster 6: binary/latched flag. This grouping informs feature pruning and representative-channel selection.

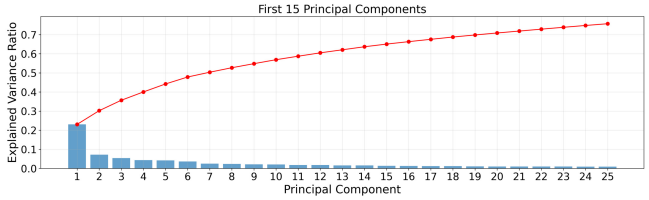

Figure 5. Explained variance (bars) for the first 25 PCs and cumulative variance (red). PC1 ≈25–26%; elbow near 5–7 PCs.

separates segments more clearly and highlights bursts/phase changes (Fig. 8). These observations motivate compact features (e.g., first k PCs, energy ratios) and time-localized analysis for downstream PHM models.

7. ANOMALY DETECTION

7.1. Supervised Fault Identification - Random Forest

Motivation. We adopt a Random Forest (RF) as a strong, low-bias baseline for multiclass anomaly detection. RFs handle high-dimensional, heterogeneous sensor channels without feature scaling, capture nonlinear interactions, offer calibrated class probabilities, and are robust when anomaly examples are relatively scarce.

Method. Raw multivariate signals are segmented into sliding windows of 3 s with a 1 s stride. Each window is flattened into a fixed-length feature vector (no normalization) and labeled with one of four classes: *Normal*, *Imbalance*, *Severe-Imbalance*, *Crack*. We train a multiclass RF with `n_estimators=100`, `max_depth=8`, `min_samples_leaf=10`, `max_features=sqrt`, `class_weight=balanced`, `bootstrap=True`. The depth/leaf constraints regularize the trees and speed up training; `class_weight=balanced` compensates for class imbalance; `max_features=sqrt` is the standard choice for decorrelated trees. At inference, the model outputs per-window class probabilities; we aggregate to flight-level via majority vote when needed.

Evaluation protocol. Because the dataset currently contains few *Crack* flights, we use a *10× Repeated Stratified Group Holdout*: in each run we select ≈30% of *flights per class* as the test set (ensuring all four classes are present), train on the remaining flights, and report window-level metrics. Grouping by flight prevents train/test leakage from overlapping windows of the same flight.

Results. Over 10 runs, **Accuracy** $= 0.986 \pm 0.014$, **Macro-F1** $= 0.986 \pm 0.015$, **Weighted-F1** $= 0.986 \pm 0.014$,

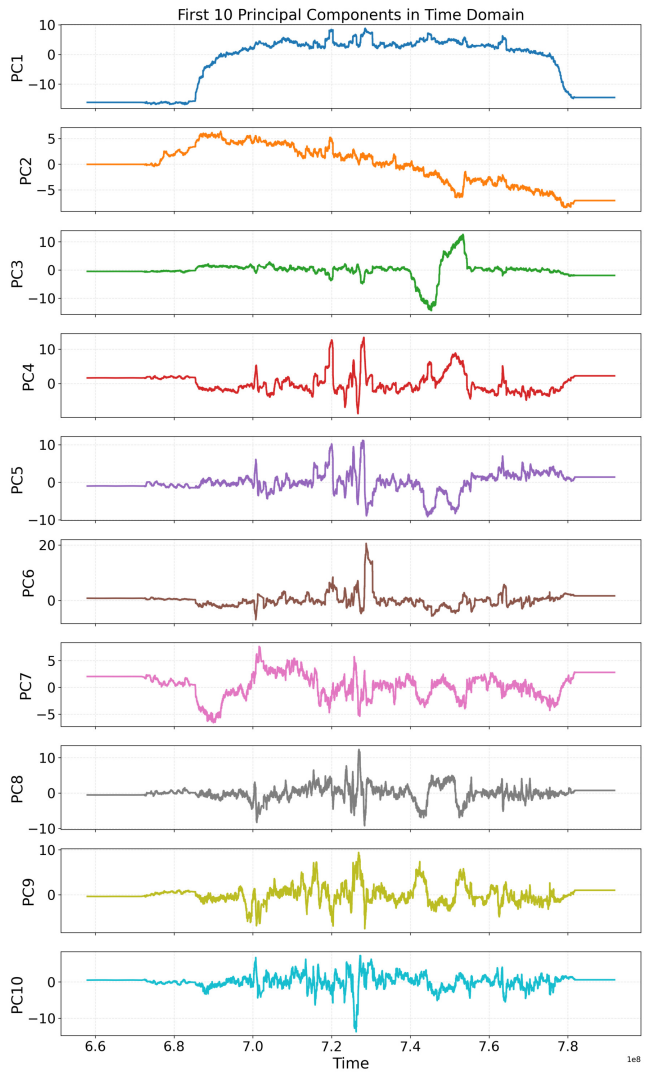

Figure 6. First ten principal–component scores versus time for one flight (after alignment and z-scaling). PC1 captures the slow flight envelope (takeoff–cruise–landing), PC2–PC3 emphasize related but distinct maneuver-related transients (one broader trend and one sharper excursion), PC4 isolates a brief step-like event, and PCs 5–10 mainly reflect higher-frequency/vibration content. (Note: PCA signs are arbitrary.)

Residual errors are dominated by occasional *Imbalance* windows predicted as *Normal*; importantly, *Crack* windows are not confused with other classes in this dataset.

Implications and limitations. These results indicate that the RF readily separates severe faults (*Crack*) from other conditions and largely distinguishes *Severe-Imbalance*/*Imbalance* from *Normal*. The remaining confusion between mild imbalance and normal behavior suggests adding more diverse imbalance examples and/or complementary features (e.g., band-power of vibration, commanded–measured residuals) could further reduce false negatives.

Table 2. Pooled 4×4 confusion matrix (rows: true; columns: predicted).

	Pred: Normal	Pred: Imbalance	Pred: Severe Imbalance	Pred: Crack
True: Normal	2994	2	1	0
True: Imbalance	71	913	0	0
True: Severe Imbalance	8	0	1014	0
True: Crack	0	0	0	897

Table 3. Confusion matrix with standard notation (TN/FP/FN/TP).

	Pred: Normal	**Pred: Anomaly**
True: Normal	TN = 2149	FP = 839
True: Anomaly	FN = 672	TP = 2278

7.2. One-Class Anomaly Detection — PCA Reconstruction Error

Motivation. To detect previously unseen faults without labels, we model *normal* behavior only: if a window cannot be well reconstructed from a low-dimensional subspace, it is flagged as anomalous. We use PCA on 3 s windows (1 s stride), fit on train-normal windows, and classify by reconstruction MSE with a threshold set from the train-normal distribution (99.5th percentile). In this section we use 64 principal components.

Results. Under the same 10× Repeated Stratified Group Holdout (flight-level grouping; all four classes present in each test), the PCA detector achieves **Accuracy** $= 0.745 \pm 0.034$, **Macro-F1** $= 0.744 \pm 0.033$, **Weighted-F1** $= 0.744 \pm 0.034$, The pooled 2×2 confusion matrix (rows true: Normal/Anomaly; columns predicted: Normal/Anomaly) is

Limitations. PCA is linear and uses a single global threshold on flattened windows; subtle, localized, or phase-shifted faults can be under-detected, while benign transients inflate false positives. As expected given the absence of label supervision, this unsupervised approach yields lower accuracy than the supervised Random Forest classifier discussed earlier. In future work, we will explore richer self-supervised models (e.g., autoencoders/forecasting transformers, frequency-band features, and per-flight calibration) to reduce FPR while improving recall on subtle anomalies.

8. DATASET SHARING

We provide access to the complete suite of raw and synchronized datasets. The raw dataset includes logs from the main

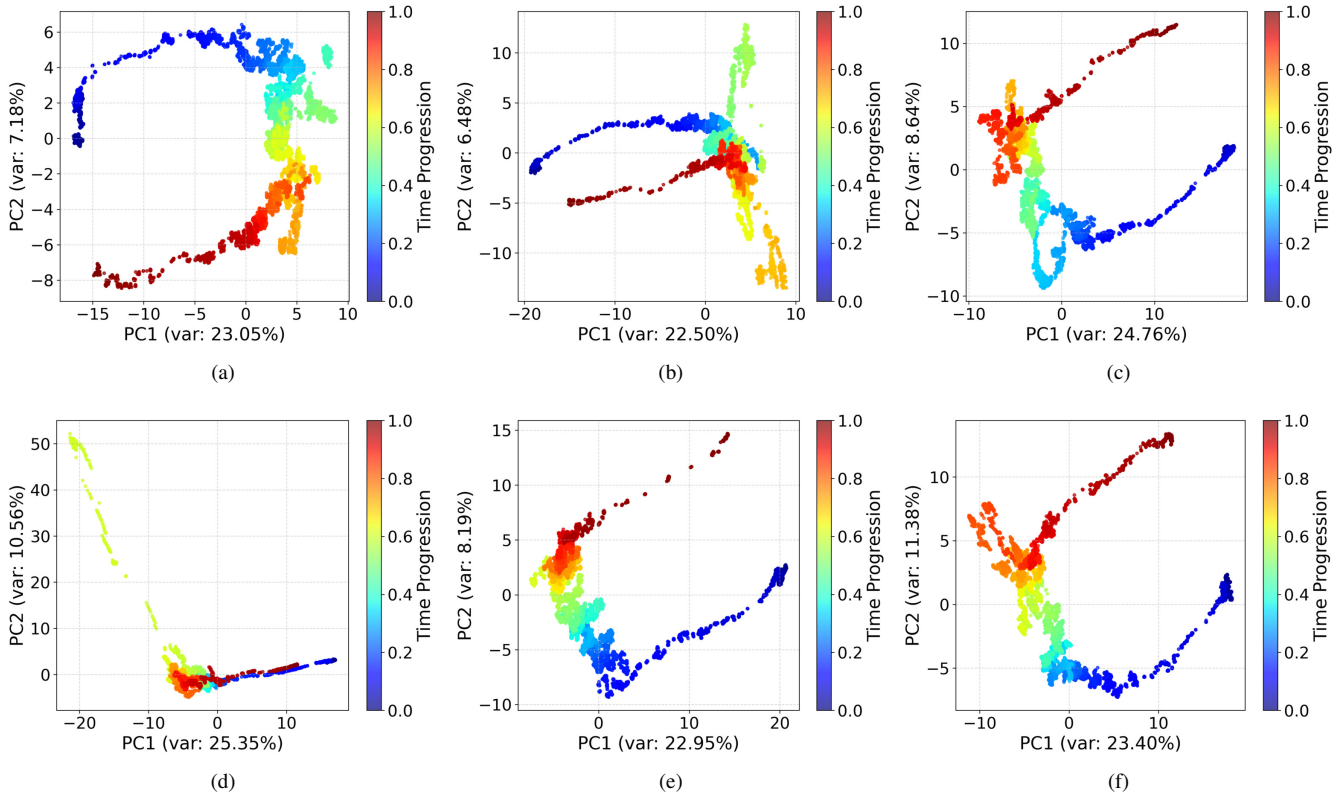

Figure 7. PC1–PC2 projections for six flights; points are colored by normalized time progression (blue→red). Top row (a–c): **Normal** flights trace smooth loops/arcs that align with flight phases (takeoff–maneuvering–landing). Bottom row (d–f): **Anomalous** flights—(d) Imbalance, (e) Severe Imbalance, (f) Cracked propeller—show broader spreads, sharper turns, and off-manifold excursions, indicating higher variability in the dominant modes. Axis annotations report the variance explained by PC1/PC2 in each case (PC1 ∼22–25%, PC2 ∼6–12%).

flight controller, auxiliary controller (vibration and temperature data), and the motion capture system, along with associated flight videos. The synchronized datasets contain timestamp-aligned logs that concatenate all processed logs into a single unified file. This dataset aggregates all relevant sensor measurements, including vibration, temperature, motion capture position, motor RPM, and battery health metrics. To improve usability and reduce dimensionality, we apply feature selection based on domain knowledge of UAV health monitoring. Redundant or low-variance variables are eliminated using code-based preprocessing pipelines, ensuring the dataset remains focused and computationally efficient.

This synchronized dataset serves as a recommended starting point for researchers interested in developing and evaluating prognostics and health management (PHM) algorithms without the overhead of extensive data preparation. By providing a clean, time-aligned, and feature-selected dataset, we enable researchers to focus on algorithm design and analysis. We also provide accompanying code snippets and documentation to facilitate dataset exploration, visualization, and baseline modeling.

The dataset is publicly available at: `https://github.com/ailab-utokyo/ut-drone-dataset`

9. CONCLUSION

This work presents a modular indoor UAV platform for PHM with rigid-mount vibration and dual-mode temperature sensing alongside standard telemetry. After cleaning and synchronization, we obtain time-aligned datasets suitable for anomaly detection and fault identification. Exploratory analyses and two reference models (supervised Random Forest; one-class PCA) illustrate utility. We release the dataset, synchronized videos, and code to accelerate research and collaboration.

Limitations & future work. The current release is indoor and single-platform with a small, imbalanced fault set (e.g., only two crack flights); induced faults do not capture gradual degradation. Next, we will (i) expand flight hours and fault coverage—including run-to-failure sequences—and (ii) validate on additional airframes and indoor/outdoor environments to assess generalization.

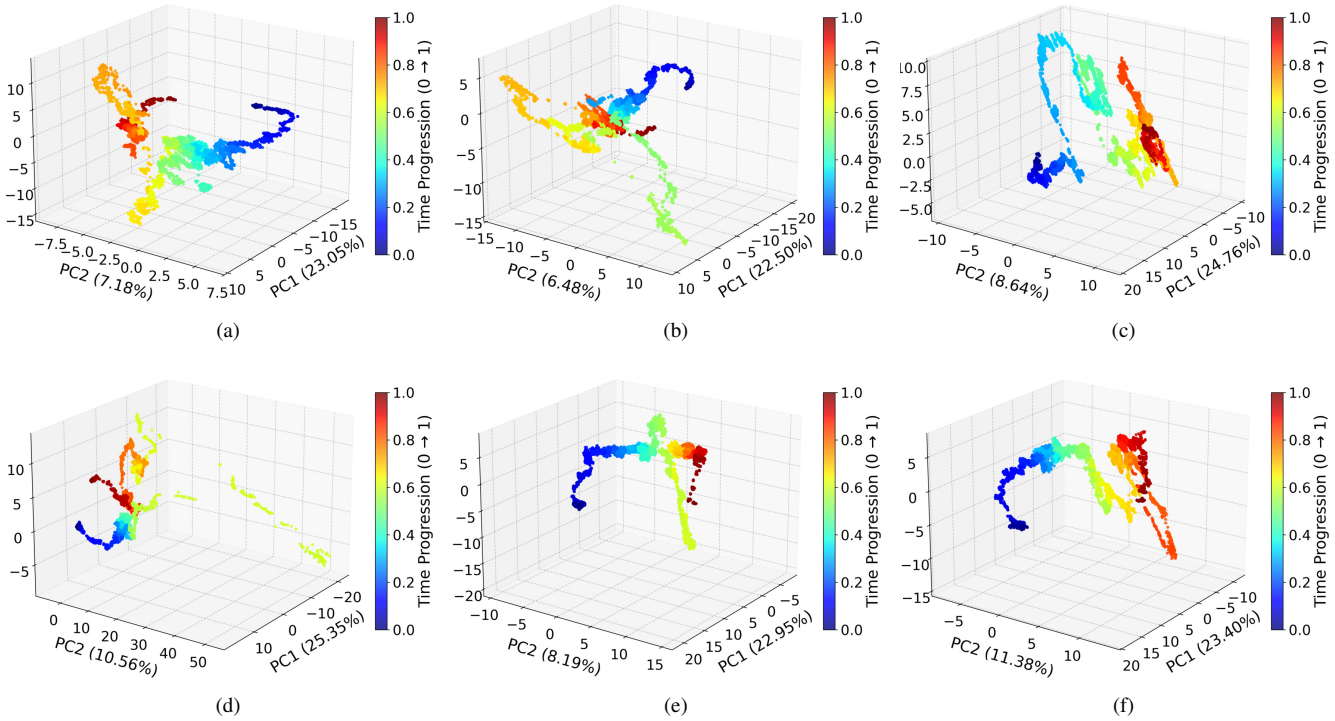

Figure 8. PC1–PC2–PC3 trajectories for six flights; points are colored by normalized time progression (blue→red). Top row (a–c): **Normal** flights trace smooth, low-curvature loops with limited depth along PC3, consistent with a stable flight envelope. Bottom row (d–f): **Anomalous** flights—(d) Imbalance, (e) Severe Imbalance, (f) Cracked propeller—exhibit thicker, more fragmented paths and larger spread in the PC3 dimension, indicating additional variability and mode excitation. Axis labels report the per-flight variance explained by PC1/PC2 (typically ∼22–25% and ∼6–12%); the added PC3 dimension reveals separations that are less visible in 2D, especially for the faulted cases.

ACKNOWLEDGMENT

This work has been supported by JSPS International Joint Research Program JPJSJRP20221501.

REFERENCES

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 93–104, 2000. 10.1145/342009.335388.

Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W. Achtelik, and Roland Siegwart. The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016. 10.1177/0278364915620033.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15:1–15:58, 2009. 10.1145/1541880.1541882.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008. 10.1109/ICDM.2008.17.

Tong Qin, Peiliang Li, and Shaojie Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018. 10.1109/TRO.2018.2853729.

Abhinav Saxena and Kai Goebel. Turbofan engine degradation simulation data set. NASA Ames Prognostics Data Repository, 2008. URL `https://data.nasa.gov/dataset/cmapss-jet-engine-simulated-data`.

Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alexander J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001. 10.1162/089976601750264965.

Takehisa Yairi, Naoya Takeishi, Tetsuo Oda, Yuta Nakajima, Naoki Nishimura, and Noboru Takata. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction. *IEEE Transactions on Aerospace and Electronic Systems*, 53(3):1384–1401, 2017. 10.1109/TAES.2017.2679680.