

Data-Driven Detection of Leaking Valves in Air Handling Units

Valve leakage in Air Handling Units (AHUs) silently degrades building energy efficiency and thermal comfort, yet it is hard to detect because compensating controls mask the fault. We study a data-driven approach that uses standard AHU telemetry to detect leakage robustly and in (near) real time.

We designed 80 controlled experiments on a full-scale laboratory AHU rig to simulate normal vs. leakage conditions under varying leakage rates, fan speeds, and water temperature profiles (1 Hz sampling; 20 min per run), yielding a labelled dataset of ~96k rows. Sensors included water/air temperatures, flow, and differential pressure (later excluded for deployability). We constructed two pipelines around gradient-boosted trees (XGBoost): (i) physically informed manual features (coil ΔT , rolling statistics, short lags), and (ii) automated feature generation (AutoFeat) to capture non-linear interactions. We also built a preprocessing bridge to align real operational logs to the lab schema to enable field validation despite limited metadata.

On unseen lab runs, the manual-feature XGBoost reached **97.41% accuracy** (ROC-AUC 0.994); the AutoFeat variant reached **99.0% accuracy** (ROC-AUC 0.998). To address the data drift between controlled lab data and real-world building data (domain shift), we injected 1,000 high-confidence field samples into training and retrained; the manual-feature model then achieved **~99% accuracy** with **ROC-AUC 0.9987** and tight cross-validation stability. Qualitative checks against coil temperature differentials (traditional leakage estimation method) corroborated predicted leakage patterns in the absence of ground-truth labels.

We implemented a lightweight inference stack—LabVIEW logging → Python watchdog → shared preprocessing/features → XGBoost inference → Streamlit dashboard—that produces operator-visible alerts with **<2** send-to-end latency on standard hardware.

The study shows that: (1) a small, interpretable feature set tied to heat-exchange physics is sufficient for reliable leakage detection; (2) a lightweight adaptation step can bridge lab-trained models to real-world building data without privileged control signals; and (3) a deployable, vendor-agnostic pipeline can be realized with commodity sensors and open-source tooling. Unlike previous HVAC fault detection studies, this work specifically targets **valve leakage detection**, a fault type that has received little to no prior attention, providing a practical path toward predictive maintenance with measurable energy and sustainability impact, while highlighting the need for larger, domain-labelled field datasets to enable large-scale deployment.