PREDICTIVE ANOMALY DETECTION IN NGAFID DATA

Anusha Seshadri and Iyashi Pal Advisor: Travis Desell 2025

RIT Software Engineering Rochester Institute of Technology

Abstract

Early detection of mechanical anomalies in general aviation is essential for improving flight safety and reducing unexpected maintenance events. Traditional point-wise anomaly detectors often struggle with NGAFID to capture the gradual cross-sensor patterns that signal emerging failures.

We present a two-stage hybrid subsequence-level anomaly detection framework for NGAFID flight data. In Stage 1, a sliding-window One-Class SVM, trained only on post-maintenance ("healthy") data, identifies abnormal patterns in pre-maintenance flights. In Stage 2, these flagged segments are analyzed using EXAMM, a neuroevolutionary system that evolves RNN architectures (LSTM, GRU, Delta-RNN) to model temporal dynamics, isolate contributing sensors, and provide localized explanations.

Results show that the OC-SVM + EXAMM pipeline reduces false positives, increases early-warning lead time, and provides sensor-level explanations via attention maps and feature rankings. This work contributes a scalable, interpretable, and aircraft-adaptive framework for predictive maintenance in aviation.

Introduction

General aviation aircraft generate large volumes of multivariate, time-varying sensor data, yet mechanical issues often emerge slowly and without clear labels. Flight crews and maintenance teams rely heavily on early detection, but gradual, cross-sensor changes are difficult to recognize manually, and traditional point-based methods rarely capture the subtle patterns that precede failures.

NGAFID adds further complexity: each flight differs in length, phase behavior, and sensor dynamics, and true anomalies are rarely documented. This creates a practical need for data-driven systems that can learn normal flight behavior directly from unlabeled, highly variable time-series data and identify deviations early enough to support proactive maintenance.

The purpose of this study is to investigate whether combining statistical and learning-based perspectives can improve the reliability and interpretability of anomaly detection for real-world aviation data. Our work is motivated by the operational importance of reducing false alarms, providing actionable explanations, and developing a framework that adapts to different aircraft and evolving flight patterns.

Methodology

We designed a multi-stage **anomaly detection** pipeline to identify early, subsequence-level deviations in flight telemetry from the National General Aviation Flight Information Database (NGAFID).

1. **Data Preparation and Subsequence Generation**
   NGAFID flight logs were cleaned, aligned, and segmented into fixed-length sliding windows to capture local flight behavior instead of treating each flight as a single long sequence.
   Each subsequence corresponded to takeoff → climb → cruise → descent phases.

2. **Baseline Detection Using One-Class SVM**
   A One-Class Support Vector Machine (OC-SVM) trained only on post-maintenance data learned the distribution of normal operating conditions.
   OC-SVM outputs: anomaly score + binary normal/outlier flag.

3. **Neuroevolutionary EXAMM Forecasting**
   Flagged windows were analyzed using EXAMM (Evolutionary eXplainable Artificial Multiobjective Model), a neuroevolution framework that evolves recurrent neural networks (LSTM, GRU, Delta-RNN).
   Models were evolved for t+1 forecasting, predicting the next timestep for all sensors.
   The prediction error vector for each window served as a richer indicator of abnormal behavior.

4. **DBSCAN Prototype Box Construction for Explainability**
   To transform prediction-error windows into interpretable anomaly categories:
   DBSCAN - Each cluster was converted into a prototype box (centroid ± percentile half-width), representing normal variation.

5. **Noise-Window Diagnostics**
   DBSCAN "noise points" (low-density windows) were further analyzed using:
   - z-score heatmaps
   - per-sensor deviation rankings
   - window-level error traces

6. **External Validation Using NAB Dataset**
   Because NGAFID lacks labeled anomaly timestamps, we evaluated our anomaly-scoring method on the Numenta Anomaly Benchmark (NAB).
   Metrics included ROC-AUC, PR-AUC, F1, precision, and recall.

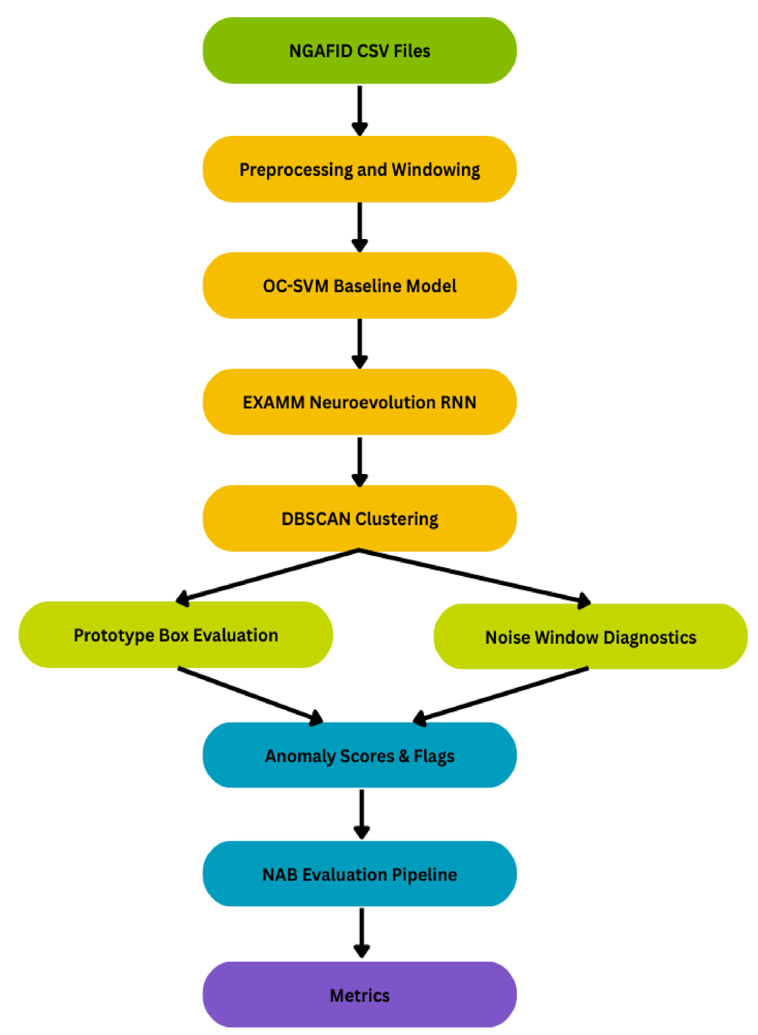


Fig. 1

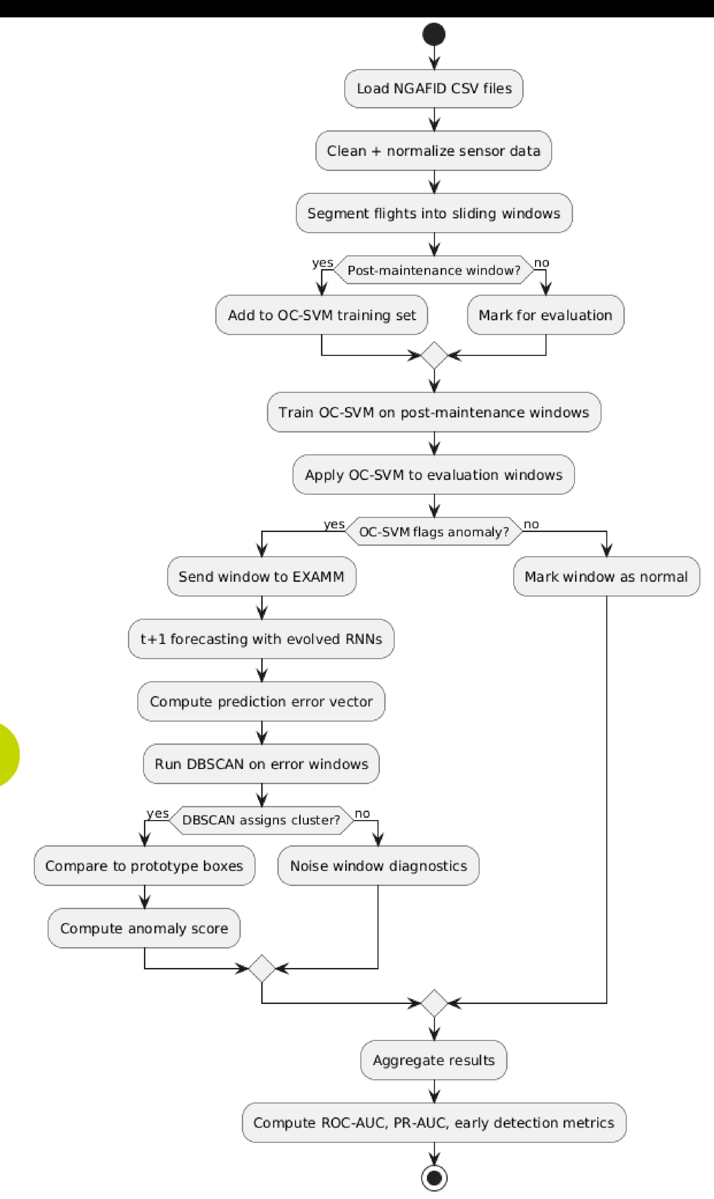


Fig. 2

Fig. 1 is the overall pipeline diagram and Fig. 2 is the architecture flowchart

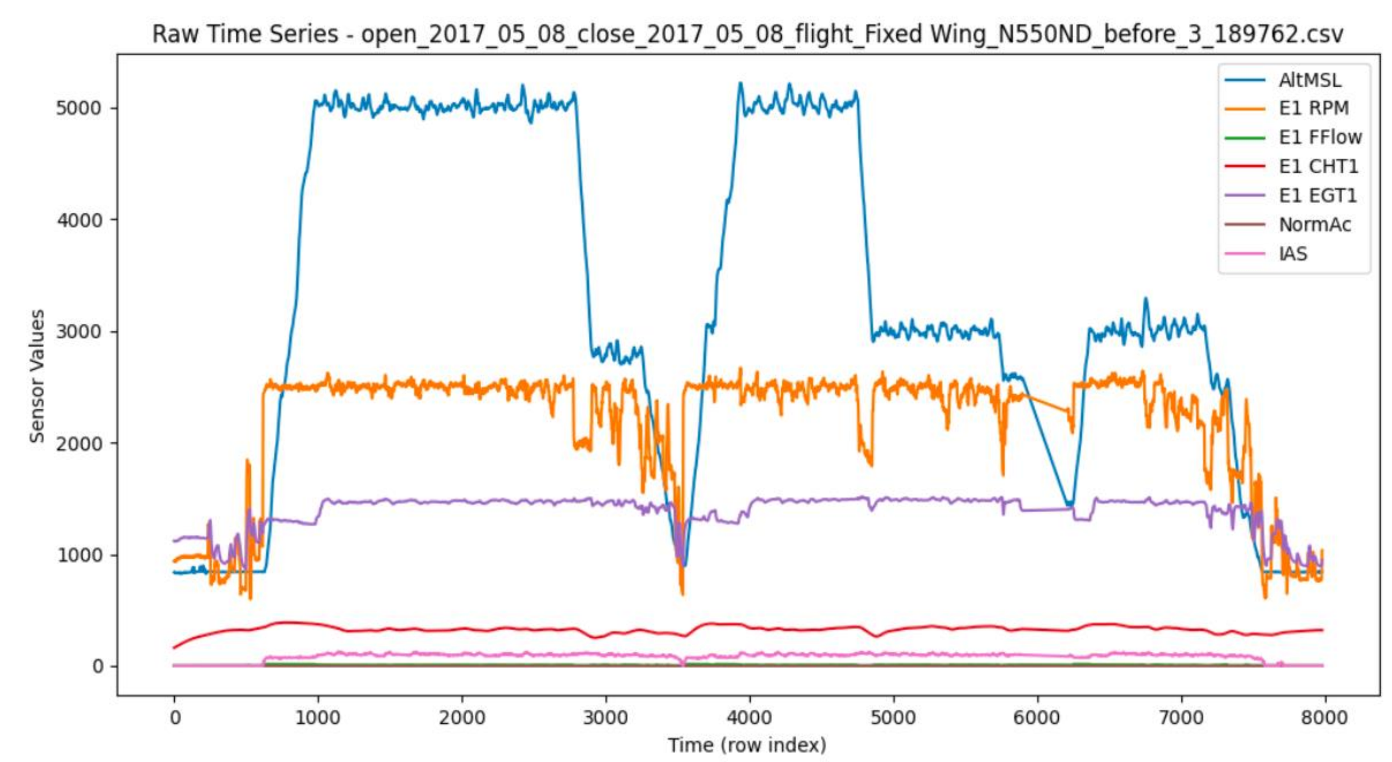


Fig. 3

Raw sensor readings for a pre-maintenance NGAFID flight subsequence, showing altitude, engine RPM, fuel flow, CHT, EGT, normal acceleration, and airspeed across the flight. Distinct climb, cruise, and descent phases are visible, along with irregular fluctuations, especially in RPM and fuel flow, that may reflect early signs of mechanical degradation.

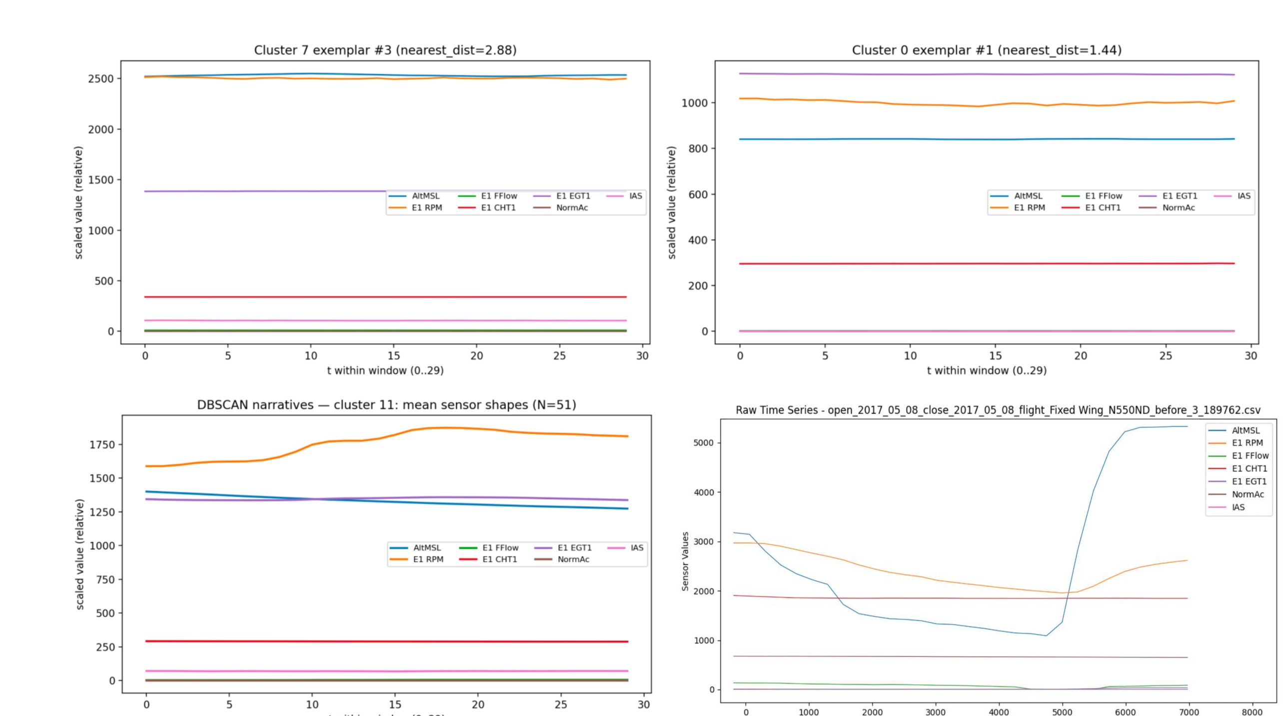


Fig. 4

Representative DBSCAN clusters showing typical subsequence shapes learned from flight data. Top panels display exemplar windows from two clusters, while the bottom-left shows the mean shape for a larger cluster (N=51). These stable patterns illustrate normal sensor behavior across windows. The bottom-right raw plot shows the corresponding full-flight context. Together, these visualizations highlight how clustering captures consistent flight dynamics and supports prototype-based anomaly scoring.

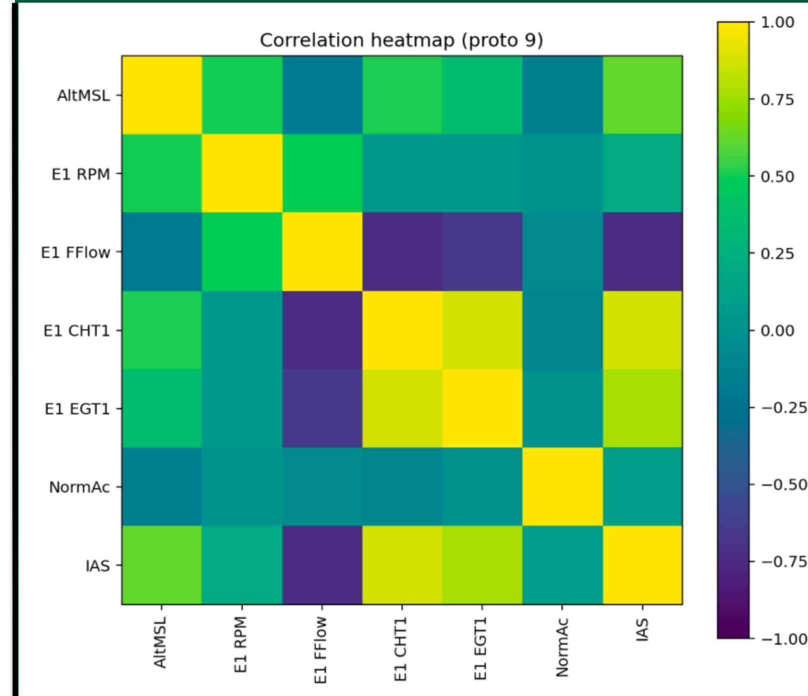


Fig. 5

Fig. 5 correlation heatmap of EXAMM prediction-error features. Strong cross-sensor relationships indicate how early anomalies propagate across multiple instruments (e.g., RPM, altitude, EGT), supporting the need for multivariate subsequence modeling.

It signifies that early failures show coordinated deviations, validating why single-sensor methods fail.

Fig. 6 DBSCAN clustering of prediction-error windows. Colored clusters represent normal behavior categories, while points labeled as noise (−1) form clear outliers. These noise windows align with known or suspected anomaly periods.

DBSCAN successfully isolates true anomalies as low-density, structurally different windows.

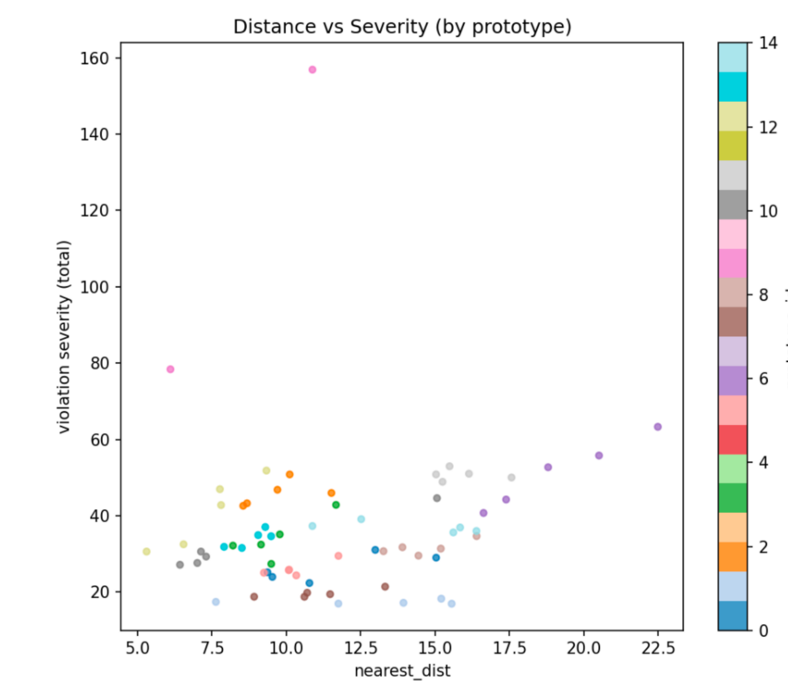
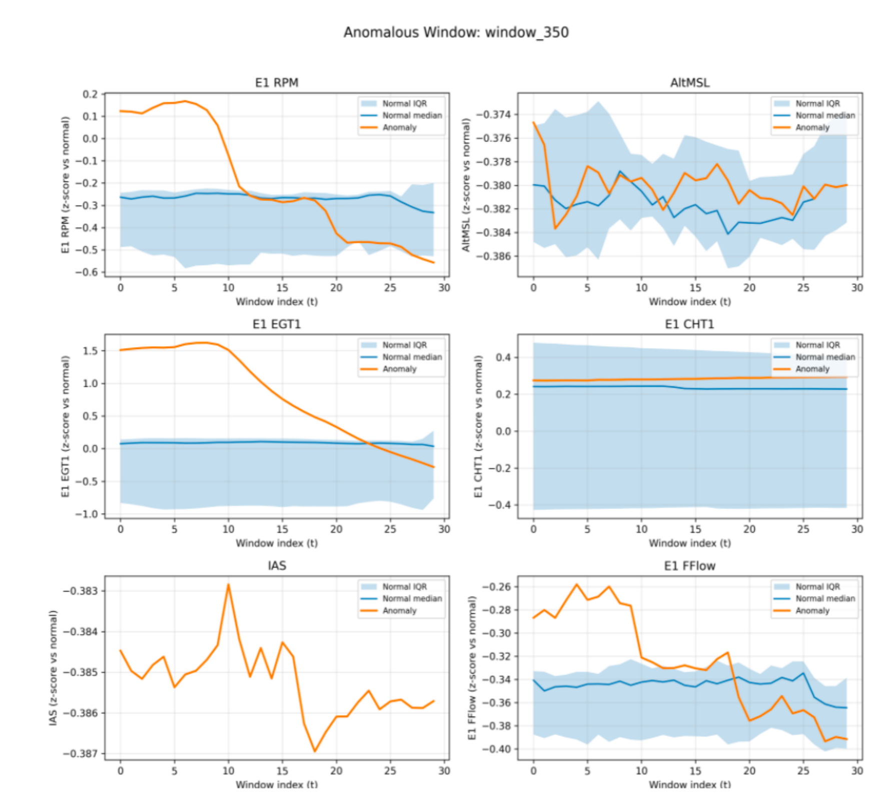


Fig. 6



Results

ROC-AUC ≈ 0.69 and PR-AUC ≈ 0.46
→ Strong for unsupervised anomaly detection on noisy flight data.

2. The system can be tuned for low false alarms OR early detection
   - High threshold → Very high precision (0.86)
   - Low threshold → Very high recall (0.76)

3. Percentile thresholds show stable recall (~0.41) across settings
This means the anomaly signal is consistent.

Conclusions

- OC-SVM + EXAMM successfully identifies early subsequence-level anomalies in NGAFID flights.
- Prediction-error windows provide clearer temporal and sensor-specific signals than pointwise methods.
- DBSCAN prototype boxes give simple, interpretable categories for normal vs. abnormal behavior.
- Noise windows consistently show strong deviations → strong indicator of true anomalies.
- NAB benchmarking shows competitive ROC-AUC/PR-AUC for a fully unsupervised method.
- System can be tuned for high precision (fewer false alarms) or high recall (early detection).

Overall, a scalable, interpretable framework for early anomaly detection in aviation.

References

[1] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," arXiv preprint, 2024.

[2] M. Pietroń, D. Żurek, K. Faber, and R. Corizzo, "AD-NEv: A Scalable Multilevel Neuroevolution Framework for Multivariate Anomaly Detection," IEEE Transactions on Neural Networks and Learning Systems, 2024.

[3] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust Anomaly Detection for Multivariate Time Series Through Stochastic Recurrent Neural Networks," Proceedings of ACM SIGKDD, 2019.

[4] A. Barbado, Ó. Corcho, and R. Benjamins, "Rule Extraction in Unsupervised Anomaly Detection for Model Explainability: Application to One-Class SVM," Expert Systems with Applications, vol. 189, 2022.

[5] A. S. Thakur, A. B. Awari, Z. Lyu, and T. Desell, "Efficient Neuroevolution Using Island Repopulation and Simplex Hyperparameter Optimization," IEEE Symposium Series on Computational Intelligence, 2023.