

A Hybrid Two-Stage Architecture for Anomaly Detection in LEO Satellite Orbit

Matteo Soldà

Department of Mathematics,
University of Padova, Padova, Italy
matteo.solda.1@studenti.unipd.it

Simone Soderi

IMT School for Advanced Studies, Lucca, Italy
Department of Mathematics,
University of Padova, Padova, Italy
simone.soderi@unipd.it

Mauro Conti

Department of Mathematics,
University of Padova, Padova, Italy
Örebro University, Örebro, Sweden
mauro.conti@unipd.it

Abstract—The congestion of Low Earth Orbit (LEO) demands robust Space Situational Awareness (SSA) to detect uncooperative satellite maneuvers. However, traditional anomaly detection models struggle to differentiate propulsive actions from space weather perturbations, yielding severe false-positive rates during periods of high solar activity. To overcome this, we propose a novel two-stage hybrid architecture for Near Real-Time (NRT) maneuver detection. First, a Bidirectional Long Short-Term Memory (Bi-LSTM) AutoEncoder (AE) extracts temporal reconstruction errors from sequential orbital telemetry. Subsequently, a Random Forest (RF) classifier fuses these deep residuals with contextual space weather and orbital kinematic features. Validated through a strict chronological split, trained during a solar minimum and tested during a solar maximum, the framework effectively decouples active maneuvers from atmospheric drag. The proposed system achieved an operational F1-Score of 0.9838, demonstrating perfect recall by capturing all historical maneuvers, while suppressing environmental noise.

Index Terms—Bi-LSTM, Random Forest, Time-Series Analysis

I. INTRODUCTION

The unprecedented proliferation of Low Earth Orbit (LEO) satellites, largely driven by the rapid deployment of commercial mega-constellations, critically challenges current Space Situational Awareness (SSA) frameworks. In this increasingly congested environment, the prompt detection of orbital anomalies is essential not only to prevent catastrophic collisions, but also to identify uncooperative or malevolent operations, such as intentional proximity eavesdropping (Figure 1) and unauthorized maneuvers. Traditional physics-based anomaly detection systems, which rely heavily on static operational thresholds and the Extended Kalman Filter (EKF), increasingly struggle to process highly non-linear orbital telemetry. Furthermore, these conventional models are exceptionally vulnerable to space weather perturbations. Rapid fluctuations in atmospheric density and drag, typical during solar maximums, often overwhelm linear predictive filters, yielding severe false-positive rates and degrading overall situational awareness.

Motivation. While deep sequence models like Bidirectional Long Short-Term Memory (Bi-LSTM) networks excel at capturing temporal dependencies in telemetry, relying solely on reconstruction errors, fail to decouple active maneuvers from natural environmental fluctuations. To overcome this, we

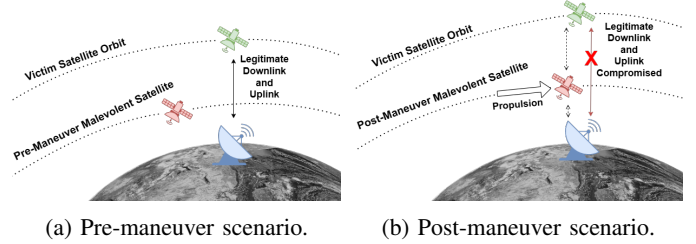


Fig. 1: Comparison between nominal (a) and malicious (b) satellite communication scenarios. In the pre-maneuver phase, the legitimate downlink and uplink operate normally. Post-maneuver, a malevolent satellite utilizes propulsion to intercept or disrupt the communication, highlighting the critical need for prompt anomaly detection to ensure SSA.

propose a hybrid two-stage framework: a Bi-LSTM AutoEncoder (AE) extracts temporal reconstruction deviations, and a Random Forest (RF) classifier fuses these residuals with multi-modal physical context (kinematics and space weather indices). This contextual fusion, refined by a 1D morphological filter, effectively suppresses weather-induced false alarms. The system is rigorously validated using Space-Track and International Laser Ranging Service (ILRS) data under a zero-trust chronological split.

Contribution.

- A hybrid two-stage architecture combining a Bi-LSTM AE with a RF classifier for anomaly detection in orbital telemetry improved by an environmental context-aware feature engineering strategy to robustly decouple atmospheric drag from propulsive accelerations with the use of the F10.7_{81_OBS}.
- A post-processing morphological filter and event-based evaluation methodology aligned with SSA operational requirements.
- A chronological evaluation across different solar activity phases, demonstrating strong robustness and low false alarm rates.

Organization. The remainder of the paper is organized as follows: Section II reviews related work, Section III provides the background needed to understand the framework,

Section IV explains the methodology, Section V describes the experimental setup, Section VI presents the experimental results and finally, Section VII summarizes the outcomes of this framework with the conclusions.

II. RELATED WORK

Spacecraft anomaly detection has evolved into advanced data-driven frameworks, comprehensively reviewed in [1] and evaluated via standardized platforms (e.g., OPS-SAT [2], [3]) or specific subsystem analyses [4].

A. Unsupervised and Traditional Machine Learning (ML)

Due to the scarcity of labeled maneuver data, unsupervised and adaptive methods remain crucial. Recent methodologies leverage frequent pattern mining combined with clustering [5], optimal proposal particle filters tailored for Two Line Elements (TLE) data [6], and adaptive algorithms for shifting environments [7]. To enhance operational generalization and interpretability, model-agnostic meta-learning [8] and Explainable AI (XAI) frameworks [9] have also been integrated into aerospace pipelines.

B. Deep Sequence Models

The sequential nature of orbital telemetry naturally aligns with Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) networks. Foundational LSTM architectures with dynamic thresholding [10] have been extensively modified to address telemetry class imbalance via Bayesian formulations [11] and subsystem-specific sequence-to-sequence modeling [12]. In the context of SSA, LSTMs have been adapted for catalog maintenance [13] and unsupervised LEO maneuver detection utilizing Bi-LSTM AEs trained on space debris ballistic baselines [14].

C. Transformers, Generative Models, and Hybrid Architectures

Recent advancements favor attention mechanisms, generative models, and hybrid architectures to capture complex telemetry dynamics. Transformer-based models effectively manage long-range dependencies [15], irregular sampling rates [16], and P-LEO constellation graph dynamics [17]. Furthermore, encoder-decoder Generative Adversarial Networks (GANs) [18], causal structure learning [19], synthetic anomaly injection [20], and hybrid data-driven frameworks [21] have expanded the state of the art. However, these models predominantly rely on endogenous data, often struggling to distinguish genuine propulsive maneuvers from environmental anomalies like space weather. Our proposed architecture bridges this gap by extending the foundational Bi-LSTM approach into a two-stage hybrid framework that explicitly fuses deep temporal residuals with external environmental context.

III. BACKGROUND

A. Bidirectional Sequence Reconstruction

Standard RNNs and unidirectional LSTMs process sequential telemetry strictly in a forward chronological direction.

While effective for causal forecasting, this approach is sub-optimal for retrospective anomaly characterization in orbital mechanics. An anomalous trajectory deviation, such as a propulsive maneuver, structurally alters the geometric context of the orbit both preceding and following the initiation event.

A Bi-LSTM addresses this limitation by processing the sequence concurrently in both forward and backward temporal directions, capturing deep, non-causal bidirectional dependencies. When integrated into an AE architecture and trained exclusively on undisturbed ballistic trajectories, the network compresses this representation into a lower-dimensional latent bottleneck. The resulting reconstruction error serves as an implicit surrogate for unmodeled physical forces, effectively isolating high-frequency propulsive accelerations from steady-state Keplerian motion.

B. Thermospheric Drag and Solar Activity Proxies

Satellites operating within LEO are primarily subjected to non-conservative perturbations dominated by atmospheric drag, parameterized by the modified ballistic coefficient (Ballistic Drag Term (BSTAR)). The critical challenge in data-driven anomaly detection stems from the high volatility of the atmospheric density.

Thermospheric density fluctuates by multiple orders of magnitude primarily due to solar Extreme Ultra Violet (EUV) radiation, which heats the upper atmosphere causing it to expand. Since direct EUV radiation cannot be continuously measured from the ground, the 10.7 cm solar radio flux ($F_{10.7}$) is globally adopted as a reliable proxy for solar-driven atmospheric heating. During periods of solar maximum, intense solar activity causes a dramatic expansion of the thermosphere. This distribution shift introduces severe non-linear perturbations into the telemetry stream, causing standalone sequential neural networks to confuse natural drag-induced orbital decay with active, uncooperative maneuvers.

C. Heterogeneous Feature Fusion via Ensemble Trees

To robustly decouple propulsive maneuvers from solar-driven drag variations, the Deep Learning (DL) temporal residuals must be explicitly contextualized using multi-modal physical variables. This task requires a classifier capable of establishing highly non-linear decision boundaries across heterogeneous data streams without suffering from exploding gradients or severe feature-scale imbalances.

The RF classifier is well suited to act as a second-stage tabular fusion engine. Unlike gradient-based deep architectures that require strict homogeneous normalization, ensemble tree structures can seamlessly ingest a highly disparate feature vector containing continuous dimensionless residuals, cyclic kinematic derivatives, and raw external space weather measurements. By computing independent thresholds across individual branches, the ensemble architecture learns conditional physical rules that map the structural interactions between neural network errors and solar cycles, optimizing precision while maintaining explainable feature importances.

IV. METHODOLOGY

The proposed framework implements a hybrid, two-stage sequential architecture designed to robustly detect LEO satellite maneuvers while actively mitigating environmental false alarms caused by space weather atmospheric drag fluctuations. The pipeline sequentially couples a Bi-LSTM AE for temporal residual extraction with an ensemble RF classifier that performs multi-modal data fusion.

A. Data Acquisition and Environmental Context

Historical orbital trajectories are retrieved via the `gp_history` endpoint of the Space-Track API, providing historical TLE sets and derived orbital parameters, such as inclination, eccentricity, RAAN and mean motion. To establish a rigorous ground truth for supervised calibration and validation, precise maneuver logs are extracted from the ILRS.

To counteract the known susceptibility of sequential neural networks to atmospheric density distribution shifts, the orbital dataset is synchronized with historical space weather indicators. Specifically, we integrate the observed 81-day centered solar radio flux index ($F_{10.7}$), denoted as `F10.7_81_OBS`, extracted from *CelesTrak*. This environmental parameter serves as a proxy for solar-driven thermospheric heating, allowing the subsequent classification stage to contextualize trajectory deviations.

B. Feature Engineering and Preprocessing

While the subsequent learning algorithms are established methodologies, the core predictive power of this framework relies on the specific integration of engineered physical and environmental features:

Kinematic Differentials. First-order temporal differences are computed for vital Keplerian elements to isolate high-frequency accelerations induced by propulsive forces $\Delta X_t = X_t - X_{t-1}$ where $X \in \{\text{Semi-major Axis, Inclination, Eccentricity}\}$.

Dynamic Drag Normalization. The ballistic coefficient (B^*), which represents atmospheric drag susceptibility, is transformed into a moving Z -score over a 60-day window to eliminate long-term seasonal trends while preserving transient propulsive anomalies $B_{\text{rolling},t}^* = \frac{B_t^* - \mu_{60D}(B_t^*)}{\sigma_{60D}(B_t^*) + \epsilon}$, where μ_{60D} and σ_{60D} represent the moving mean and standard deviation over a 60-day baseline, respectively, and $\epsilon = 10^{-9}$ is a stabilization constant.

Chronological Split. To enforce a strict zero-trust validation environment and simulate operational deployment, the dataset is partitioned using a hard chronological split at 2022-01-01. The training partition isolates the satellite's history during a solar minimum phase, while the test partition exposes the architecture to intense solar maximum conditions. A `StandardScaler` is fitted exclusively on the nominal training distribution to compute standardization parameters, preventing data leakage. Continuous data streams are structured into overlapping temporal sequences ($\text{stride} = 1$ for maximum overlapping) using a sliding window of size $w =$

9. Operationally, while computational inference is instantaneous, Near Real-Time (NRT) detection latency is intrinsically bounded by the $w = 9$ temporal window: assuming an average of 4 TLE updates daily, a maneuver is fully detected 2 to 3 days post-event.

C. Stage 1: Bi-LSTM AE for Temporal Feature Extraction

The first stage compresses multi-variate sequential orbital elements into a dense temporal anomaly score. Although Bi-LSTM AEs are standard, established techniques for time-series representation, their specific application here serves a targeted architectural purpose: acting strictly as a temporal residual extractor to feed the subsequent data fusion stage, rather than functioning as a standalone anomaly detector. The network utilizes a specialized bottleneck architecture with an asymmetric unidirectional decoder. The architecture is graphically described in Figure 2. The network architecture is structured as follows:

- **Encoder:** Consists of three consecutive Bi-LSTM layers with decreasing hidden units ($32 \rightarrow 16 \rightarrow 8$). This configuration processes the input sequence simultaneously in forward and backward temporal directions, capturing deep bidirectional dependencies of the ballistic trajectory.
- **Latent Bottleneck:** The final encoder representations are mapped to a lower-dimensional latent space via a fully-connected layer with a dimension of 3 ($\text{latent_dim}=3$). This dimensionality was selected following empirical grid-search, forcing the AE to map the intrinsic 3D spatial dynamics of deterministic Keplerian mechanics while aggressively discarding higher-dimensional, high-frequency sensor noise. A Repeat Vector layer replicates this latent representation w times to condition the decoder.
- **Decoder:** To enforce causal reconstruction, the decoding pipeline utilizes three consecutive Unidirectional LSTM layers with expanding capacities ($16 \rightarrow 32 \rightarrow 64$), culminating in a `TimeDistributed` dense layer that outputs the reconstructed sequence $\hat{X}_t$.

The reconstruction discrepancy is quantified via the Mean Absolute Error (MAE) across the sequence windows. To ensure scale invariance across varying orbital regimes, a rolling 15-day Z -score is applied to the raw MAE values, yielding the standardized temporal anomaly feature:

$$\text{MAE}_t = \frac{1}{w \cdot f} \sum_{i=1}^w \sum_{j=1}^f |X_{i,j} - \hat{X}_{i,j}|, \quad (1)$$

$$\text{LSTM_MAE_Z}_t = \frac{\text{MAE}_t - \mu_{15D}(\text{MAE}_t)}{\sigma_{15D}(\text{MAE}_t) + \epsilon}, \quad (2)$$

where $X_{i,j}$ and $\hat{X}_{i,j}$ represent the true and reconstructed values of the j -th feature at the i -th time step within the window, w is the window size, f denotes the total number of input features, μ_{15D} and σ_{15D} denote the moving mean and standard deviation of the MAE computed over a rolling 15-day window, and ϵ is a small constant added for numerical stability.

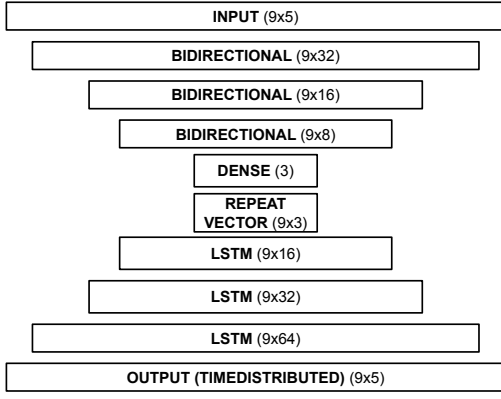


Fig. 2: Simplified architectural schema of the asymmetric Bi-LSTM AE. The network comprises a bidirectional encoder for deep temporal feature extraction and a unidirectional decoder for causal sequence reconstruction, connected by a reduced latent bottleneck (dimension 3) to filter high-frequency noise. Numbers inside the squares indicate the dimensionality of each layer. Detailed hyperparameters and layer configurations are provided in IV-C

D. Stage 2: Heterogeneous Data Fusion via RF Classification

Rather than using a rigid analytical threshold on the reconstruction error, the final decision boundary is established by a RF classifier acting as a tabular data-fusion engine. While the RF is a well-established classification algorithm, the primary contribution of this stage lies in its novel integration strategy: exploiting the ensemble trees to map the non-linear structural interactions between deep learning temporal residuals and explicitly engineered space weather context. The classifier receives a 5-dimensional heterogeneous feature vector v_t at each time step:

$$v_t = [\text{LSTM_MAE_Z}_t, \text{F10.7_81_OBS}_t, \Delta \text{Semi-major Axis}_t, \Delta \text{Inclination}_t, B_{\text{rolling_Z},t}^*] \quad (3)$$

The RF architecture is configured with 200 estimators and a maximum tree depth of 8, alongside a balanced class-weight strategy to counteract the extreme scarcity of maneuver observations. These hyperparameters were optimized via cross-validation on the training partition. A depth of 8 provided the optimal trade-off: it ensured sufficient complexity to map non-linear structural interactions between neural network errors and solar cycles, while strictly preventing overfitting on the heavily imbalanced maneuver minority class. By fusing the neural network's reconstruction residual with direct space weather tracking and kinematic derivatives, the ensemble classifier learns to contextualize the source of orbital deviations. Consequently, if an elevation in the LSTM_MAE_Z coincides with a sudden spike in the solar flux (F10.7_81_OBS) and a matching elevation in BSTAR_ROLLING_Z without a corresponding kinematic shift in inclination, the classifier effectively suppresses the false alarm, categorizing the event as an atmospheric drag perturbation. Figure 3 shows the graphical description of this two-stage architecture.

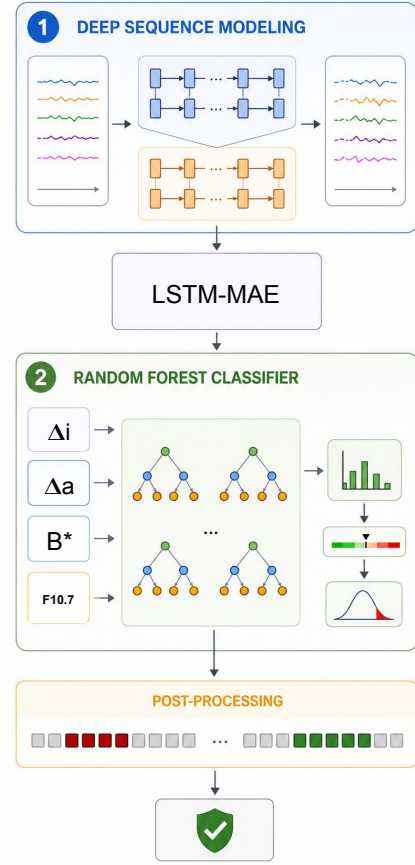


Fig. 3: Comprehensive overview of the proposed two-stage hybrid anomaly detection framework. Stage 1 extracts deep temporal residuals via the Bi-LSTM AE. Stage 2 fuses these residuals with heterogeneous physical context (kinematics and space weather) using a Random Forest classifier. The final post-processing step applies a 1D morphological filter to identify persistent maneuver-related anomalies while robustly suppressing false positives induced by atmospheric drag.

E. Training Protocol

The hybrid framework is trained sequentially. In the semi-supervised phase, the Deep Bi-LSTM AE is exposed exclusively to nominal training sequences, optimizing network weights by minimizing the Mean Squared Error (MSE) loss function via the Adam optimizer with a gradient norm constraint ($\text{clipnorm}=1.0$). An exponential learning rate decay schedule is applied, initialized at 0.001. Early stopping with a patience of 50 epochs monitors the validation loss to prevent overfitting.

In the second phase, the RF classifier is trained under a supervised regime. It maps the generated hybrid space v_t of the training split (containing both nominal records and the designated exclusion windows flanking historical maneuvers) to establish the definitive operational decision boundary. The importance of the five most important features of the RF is

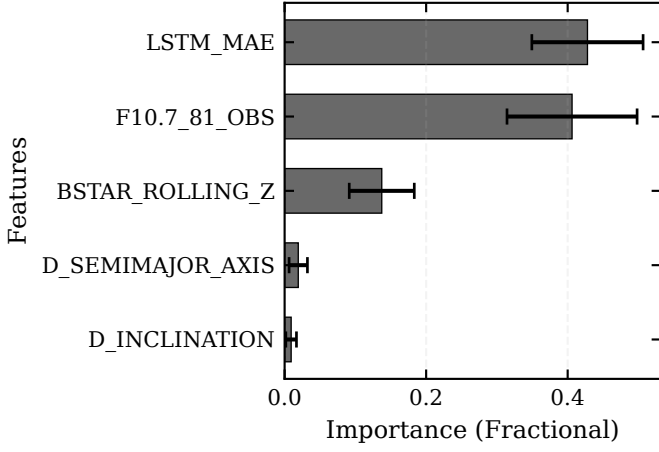


Fig. 4: Random Forest feature importance distribution. The graph illustrates the relative weight of each input feature in establishing the final classification boundary. The temporal reconstruction error (LSTM_MAE) and the solar flux index (F10.7_81_OBS) are the most discriminative features, collectively accounting for approximately 84% of the decision weight, demonstrating the necessity of environmental contextualization.

shown in Figure 4.

V. EXPERIMENTAL SETUP

A. Anomaly Detection Mechanism and Threshold

The Bi-LSTM AE outputs a reconstructed sequence for each input window. The anomaly score is defined as the MSE between the input sequences and their respective reconstructions. Under nominal conditions, the model accurately reconstructs the ballistic trajectory, yielding a low MSE. Conversely, during an orbital maneuver, the unmodeled dynamics cause a spike in the reconstruction error.

To convert the continuous MSE anomaly score into a binary classification (Nominal vs. Anomaly) without introducing data leakage, the threshold was calibrated exclusively using the nominal data distribution. In SSA operations, failing to detect an active maneuver (False Negative) poses a significantly higher operational risk than investigating a false alarm (False Positive). To enforce this conservative safety margin, the anomaly threshold (τ) was defined as the 90th percentile of the MSE distribution computed strictly on the nominal validation set:

$$\tau = P_{90}(MSE_{val}), \quad (4)$$

where any temporal sequence in the operational or test phase exhibiting an MSE strictly greater than τ is flagged as an anomaly.

B. Evaluation Metrics

The predictive capability of the proposed architecture was evaluated using standard classification metrics computed on the test set:

- **Precision:** The ratio of correctly identified anomalous sequences to the total sequences flagged as anomalies. It reflects the model’s robustness against False Positives (sensor noise mistaken for maneuvers).
- **Recall (Sensitivity):** The ratio of correctly identified anomalous sequences to the total actual anomalies present.
- **F1-Score:** The harmonic mean of Precision and Recall, providing a single metric for the model’s balance.
- **ROC-AUC:** The Area Under the Receiver Operating Characteristic Curve, used to assess the overall discriminative capability of the continuous MSE score across all possible thresholds.

C. Computing Environment

The data processing and hybrid architecture were implemented in Python using TensorFlow, Keras, and Scikit-learn, with the training phase hardware-accelerated on an NVIDIA GeForce RTX 3090 GPU.

VI. RESULTS AND DISCUSSION

The experimental evaluation was designed to rigorously test the framework’s ability to discriminate between active satellite maneuvers and environmental noise across shifting solar activity cycles. Performance was quantified through both traditional point-wise metrics and operationally relevant event-based metrics.

The comparison between our proposed framework and the ones discussed in Section II is summarized in Table I.

A. Stage 1: Standalone Bi-LSTM Performance and Vulnerabilities

The initial evaluation focused on the standalone performance of the Deep Bi-LSTM AE prior to the multi-modal fusion stage. Grid-search optimization identified the bottleneck architecture with a sliding window of $w = 9$ as the most effective temporal extractor. On the isolated nominal testing splits, this configuration achieved strong classification performance, peaking at an F1-Score of 0.9436, a Precision of 0.9592, and a Recall of 0.9285 (ROC-AUC: 0.8984).

However, to simulate a realistic SSA operational deployment, the model was subjected to a strict *Full-Life Validation* stress test. This zero-trust evaluation spanned the entirety of the satellite’s active history, deliberately crossing the boundary from a solar minimum (training phase) into an intense solar maximum (testing phase). Under these volatile environmental conditions, the standalone sequential model’s performance degraded significantly. The optimal full-life F1-Score plummeted to 0.6410. While the model maintained a high Precision (0.9693), its Recall collapsed to 0.4788, indicating that pure residual-based neural networks struggle to adapt to extreme atmospheric density shifts, ultimately missing more than half of the relevant orbital deviations. This degradation confirms the operational necessity of the second-stage environmental fusion.

TABLE I: Methodological Comparison Between the State-of-the-Art Literature and the Proposed Framework

Reference	Deep Seq. Models (LSTM/RNN)	Transformers/ Attention	TLE / Orbital Data Focus	Hybrid Fusion Pipeline	Space Weather Integration	Event-Based Post-Proc.
Hundman et al. [10]	✓	—	—	—	—	—
Chen et al. [11]	✓	—	—	—	—	—
Wang et al. [20]	✓	—	—	—	—	—
Nalepa et al. [3]	—	—	—	—	—	—
Xu et al. [21]	—	—	—	✓	—	—
Cuéllar et al. [9]	—	—	—	—	—	—
Indaco et al. [17]	—	✓	✓	—	—	—
Shorten et al. [6]	—	—	✓	—	—	—
Cipollone et al. [13]	✓	—	✓	—	—	—
Raviola [14]	✓	—	✓	—	—	—
Fejjari et al. [15]	—	✓	—	—	—	—
Dai et al. [16]	—	✓	—	—	—	—
Djerida [5]	—	—	—	—	—	—
Proposed Framework	✓	—	✓	✓	✓	✓

The symbol ✓ indicates the presence of the feature, whereas — indicates its absence.

B. Stage 2: Hybrid Data Fusion and RF Classification

To address the vulnerabilities of the standalone neural network, the temporal reconstruction error (LSTM_MAE) was standardized into a moving Z-score and fused with kinematic differentials and space weather indices (F10.7) via the RF classifier.

When evaluated using traditional point-wise predictions, the RF data-fusion engine successfully stabilized the detection baseline. The raw classification output yielded an F1-Score of 0.8206, restoring extreme sensitivity with a Recall of 0.9919, albeit with a moderate Precision of 0.6997 due to transient noise spikes. This demonstrates that the core machine learning architecture is primarily responsible for maximizing maneuver detection capability (Recall) under volatile solar conditions, independently of any operational post-processing.

C. Post-Processing and Final Operational Metrics

While the raw RF classifier maximizes sensitivity, transient environmental noise limits its point-wise precision. To align the model's output with actionable intelligence, a 1D morphological filter (implementing consecutive closing and anti-spike operations) was applied to the raw RF probabilities. This post-processing step is strictly responsible for the precision gains, filtering isolated false positives without altering the underlying learning model's predictive boundary. Furthermore, the final evaluation was conducted using an event-based strategy, granting an asymmetric forward temporal tolerance of 9 days to account for General Perturbation (GP) catalog update delays.

The application of this post-processing pipeline produced outstanding definitive operational metrics. By explicitly separating the performance gains, it is evident that the morphological filtering and event-based grouping elevated the raw F1-Score from 0.8206 to a final Event-Based F1-Score of 0.9838. The system maintained near-perfect sensitivity (Recall of 1.0000) by successfully identifying all 91 active maneuvers, while the post-processing effectively suppressed the remaining atmospheric false alarms, surging Precision from 0.6997 to

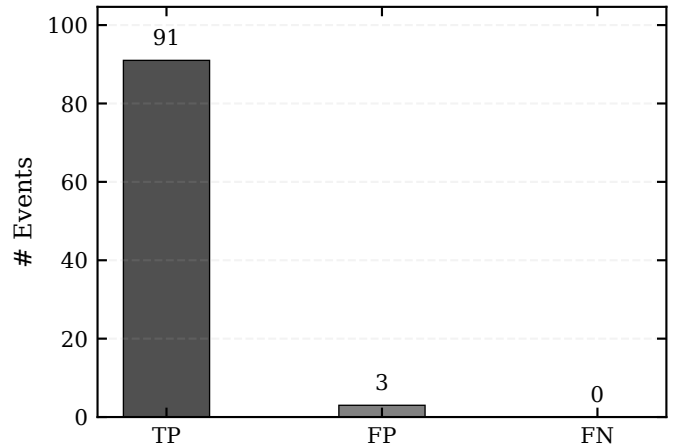


Fig. 5: Summarized event-based classification results for the Sentinel-3A satellite over the full lifecycle testing period. The hybrid framework achieved perfect recall by detecting all 91 confirmed maneuvers (True Positives) while maintaining high precision, registering only 3 False Positives despite volatile solar maximum conditions.

0.9681 (only 3 False Positives). The result is summarized in Figure 5

D. Thresholding Trade-offs

To further contextualize these results, a high-resolution threshold sweep was conducted to explore the limits of conservative detection. By enforcing an extremely strict anomaly threshold (0.01), the system could achieve absolute perfection in Precision (1.0000, False Positives = 0). However, this conservative bias severely hampered sensitivity, dropping the Recall to 0.6276 and resulting in 54 missed maneuvers (False Negatives). This comparative analysis proves that the proposed methodology, pairing hybrid data fusion with morphological smoothing, represents the optimal trade-off, maximizing situ-

ational awareness without overwhelming operators with false alerts. The results of the various tests are described in Table II in which are reported the Bi-LSTM Test F1, the Bi-LSTM Full History F1, the RF Point Wise F1 and finally the RF Event-Based F1. Each F1-score is reported along with its corresponding precision and recall. The models were tested on a subset of satellites having available and up-to-date ILRS maneuvers files, comprehending *Sentinel 3A*, *Sentinel 3B*, *Sentinel 6A* and *CryoSat 2*.

TABLE II: Performance metrics across different evaluation stages for the analyzed LEO satellites.

Satellite	Evaluation Stage	F1-Score	Precision	Recall
<i>Sentinel-3A</i>	Test LSTM	0.94	0.95	0.92
	Full LSTM	0.64	0.96	0.47
	PointWise RF	0.82	0.70	0.99
	EventBased RF	0.98	0.97	1.00
<i>Sentinel-3B</i>	Test LSTM	0.97	0.97	0.97
	Full LSTM	0.86	0.94	0.79
	PointWise RF	0.74	0.70	0.80
	EventBased RF	0.95	0.94	0.96
<i>Sentinel-6A</i>	Test LSTM	0.89	0.83	0.95
	Full LSTM	0.72	0.74	0.70
	PointWise RF	0.27	0.15	1.00
	EventBased RF	0.91	0.84	1.00
<i>CryoSat-2</i>	Test LSTM	0.95	0.97	0.94
	Full LSTM	0.66	0.95	0.50
	PointWise RF	0.76	0.76	0.77
	EventBased RF	0.86	0.78	0.97

VII. CONCLUSION

This paper presented a novel two-stage hybrid framework for NRT satellite maneuver detection in LEO. By coupling a Deep Bi-LSTM AE with a RF data-fusion engine, the proposed architecture effectively decouples active propulsive maneuvers from natural orbital decay caused by solar-driven thermospheric drag. Experimental evaluations across varying solar cycles demonstrated that while standalone sequential models suffer severe performance degradation during solar maxima, the multi-modal integration of orbital telemetry, kinematic differentials, and space weather indices ($F_{10.7}$) guarantees robust anomaly detection. Validated through a strict chronological stress test and an event-based morphological filter, the context-aware framework achieved near-perfect operational sensitivity while drastically suppressing false alarms. Because the temporal extraction and data fusion stages execute independently per satellite, the computational complexity scales linearly $\mathcal{O}(N)$ with the number of constellation nodes N . This decoupled inference architecture is highly parallelizable, enabling straightforward deployment on distributed commercial off-the-shelf infrastructure. Future work will explore the operational scaling of this approach to massive mega-constellations containing thousands of nodes, alongside its potential integration with spatial-temporal graph networks.

REFERENCES

- [1] A. Fejjari, A. Delavault, R. Camilleri, and G. Valentino, "A review of anomaly detection in spacecraft telemetry data," *Applied Sciences*, vol. 15, no. 10, p. 5653, 2025.
- [2] B. Ruszczak, K. Kotowski, D. Evans, and J. Nalepa, "The ops-sat benchmark for detecting anomalies in satellite telemetry," *Scientific Data*, vol. 12, p. 710, 2025.
- [3] J. Nalepa, M. Myller, J. Andrzejewski, P. Benecki, S. Piechaczek, and D. Kostrzewa, "Evaluating algorithms for anomaly detection in satellite telemetry data," *Acta Astronautica*, vol. 198, pp. 689–701, 2022.
- [4] A. Adam, A. Mokhtar, A. Mattar, M. Abdrahman, and S. Helmy, "A comparative analysis of anomaly detection techniques for battery telemetry data in low earth orbit remote sensing satellites," *International Journal of Prognostics and Health Management*, vol. 16, no. 2, 2025.
- [5] A. Djerida, "Unsupervised anomaly detection for satellite telemetry data using frequent pattern mining and clustering approach (fpmc)," *Advances in Space Research*, vol. 77, no. 3, pp. 3718–3731, 2026.
- [6] D. P. Shorten, M. Humphries, J. Maclean, Y. Yang, and M. Roughan, "Optimal proposal particle filters for detecting anomalies and manoeuvres from two line element data," *Acta Astronautica*, vol. 228, pp. 709–723, 2025.
- [7] L. Brancato, A. Lucchetti, F. Cadini, and M. Giglio, "Towards adaptive and robust unsupervised anomaly detection in satellite telemetry," in *Proceedings of the PHM Society Asia-Pacific Conference 2025*, 2026.
- [8] J. Li, X. Tang, W. Li, C. Liu, H. Zhang, and Q. Liu, "An anomaly detection approach for aerospace telemetry data based on model-agnostic meta-learning," *IFAC-PapersOnLine*, vol. 59, no. 20, pp. 1824–1829, 2025.
- [9] S. Cuéllar, M. Santos, F. Alonso, E. Fabregas, and G. Farias, "Explainable anomaly detection in spacecraft telemetry," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108083, 2024.
- [10] K. Hundman, V. Constantinou, M. Laporte, I. Colwell, and D. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 387–395.
- [11] J. Chen, D. Pi, Z. Wu, X. Zhao, Y. Pan, and Q. Zhang, "Imbalanced satellite telemetry data anomaly detection model based on bayesian lstm," *Acta Astronautica*, vol. 180, pp. 232–242, 2021.
- [12] W. Jin, S. Zhang, B. Sun, P. Jin, and Z. Li, "An analytical investigation of anomaly detection methods based on sequence to sequence model in satellite power subsystem," *Sensors*, vol. 22, no. 5, p. 1819, 2022.
- [13] R. Cipollone, C. Früh, and P. D. Lizia, "Lstm-based maneuver detection for resident space object catalog maintenance," *Neural Computing and Applications*, vol. 37, pp. 17 341–17 362, 2025.
- [14] E. Raviola, "An unsupervised leo satellite maneuver detection method using recurrent neural networks," Master's thesis, Politecnico di Milano, 2025, master's Thesis in Space Engineering. Advisor: Prof. Pierluigi Di Lizia, Co-advisor: Riccardo Cipollone.
- [15] A. Fejjari, A. Delavault, R. Camilleri, and G. Valentino, "Transformer-based anomaly detection for satellite telemetry data," *Acta Astronautica*, vol. 238, no. A, pp. 739–745, 2026.
- [16] A. Dai, Y. Xiao, F. Ren, H. Shen, S. Zhi, and B. Ma, "Temporal-enhanced transformer for anomaly detection in spacecraft irregular-interval telemetry," *Advances in Space Research*, vol. 77, no. 2, pp. 2356–2371, 2026.
- [17] M. Indaco and D. Guzzetti, "Transformer-based anomaly detection in p-leo constellations: A dynamic graph approach," *Acta Astronautica*, vol. 218, pp. 177–194, 2024.
- [18] Z. Xu, Z. Cheng, Q. Tang, and B. Guo, "An encoder-decoder generative adversarial network-based anomaly detection approach for satellite telemetry data," *Acta Astronautica*, vol. 213, pp. 547–558, 2023.
- [19] S. Chen, G. Jin, and X. Long, "On-orbit satellite hierarchical anomaly detection using causal structure learning," *Advances in Space Research*, vol. 75, no. 1, pp. 718–736, 2025.
- [20] Y. Wang, J. Gong, J. Zhang, and X. Han, "A deep learning anomaly detection framework for satellite telemetry with fake anomalies," *International Journal of Aerospace Engineering*, vol. 2022, p. 1676933, 2022.
- [21] Z. Xu, Z. Cheng, and B. Guo, "A hybrid data-driven framework for satellite telemetry data anomaly detection," *Acta Astronautica*, vol. 205, pp. 281–294, 2023.