

## Case Study on Detecting Anomalies in CubeSat Telemetry Using Machine Learning Approach

Sakshi Vittal\*

School of Aerospace, Transport and Management, Cranfield University: Cranfield, England.

**Abstract:** The increasing complexity of CubeSat missions and the volume of telemetry data they generate has heightened the need for advanced anomaly detection systems capable of identifying faults before they escalate into mission failures. Traditional threshold-based monitoring approaches fall short in capturing subtle, multivariate, and time-dependent anomalies inherent in satellite telemetry. This case study explores the application of machine learning (ML) techniques for anomaly detection in CubeSat telemetry, with a focus on evaluating recent research supported by publicly released benchmark datasets: OPSSAT-AD and ESA-ADB. These datasets provide real-world, labelled telemetry from operational ESA missions and have enabled systematic benchmarking of over 30 machine learning models. The study synthesises model performance across metrics such as F1-score and AUC, highlighting that temporal models like LSTMs and temporal convolutional networks (TCNs) consistently outperform classical methods in time-series tasks. Limitations in dataset continuity, anomaly sparsity, and generalisability are discussed, along with the need for explainable and resource-efficient models suitable for onboard deployment. The case concludes with a call for expanded benchmark datasets, real-time validation, and cross-disciplinary collaboration to ensure robust, interpretable, and mission-ready anomaly detection systems for future CubeSat applications.

### Table of Contents

1. Introduction.....	1
2. Traditional Vs. Machine Learning (ML) Based Anomaly Detection Approaches .....	3
3. Overview of Relevant ML Models .....	4
4. Datasets.....	5
5. Review of Existing Studies.....	6
6. Discussion .....	8
7. Conclusion.....	9
8. References .....	10
9. Conflict of Interest .....	10
10. Funding .....	10

### 1. Introduction

The aviation industry is under increasing pressure to decarbonize as part of the global response to climate change. CubeSats are a class of miniaturized satellites based on a standardized unit size of  $10 \times 10 \times 10$  cm (1U), initially developed to support low-cost access to space for academic institutions [1]. Over the last decade, CubeSats have evolved from simple educational tools into sophisticated platforms for Earth observation, scientific experiments, and commercial applications. Their modular design, low development cost, and rapid deployment capabilities have made them popular among space agencies, research institutions, and startups [2].

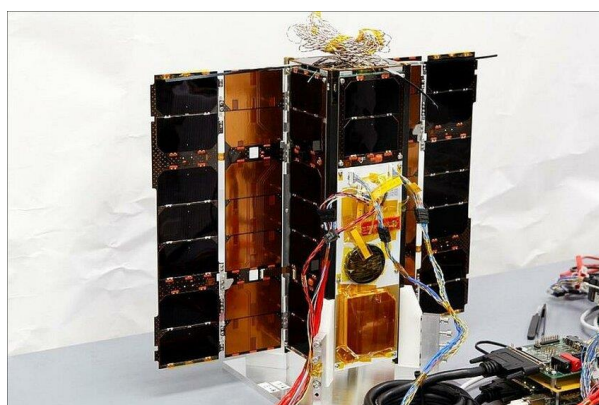
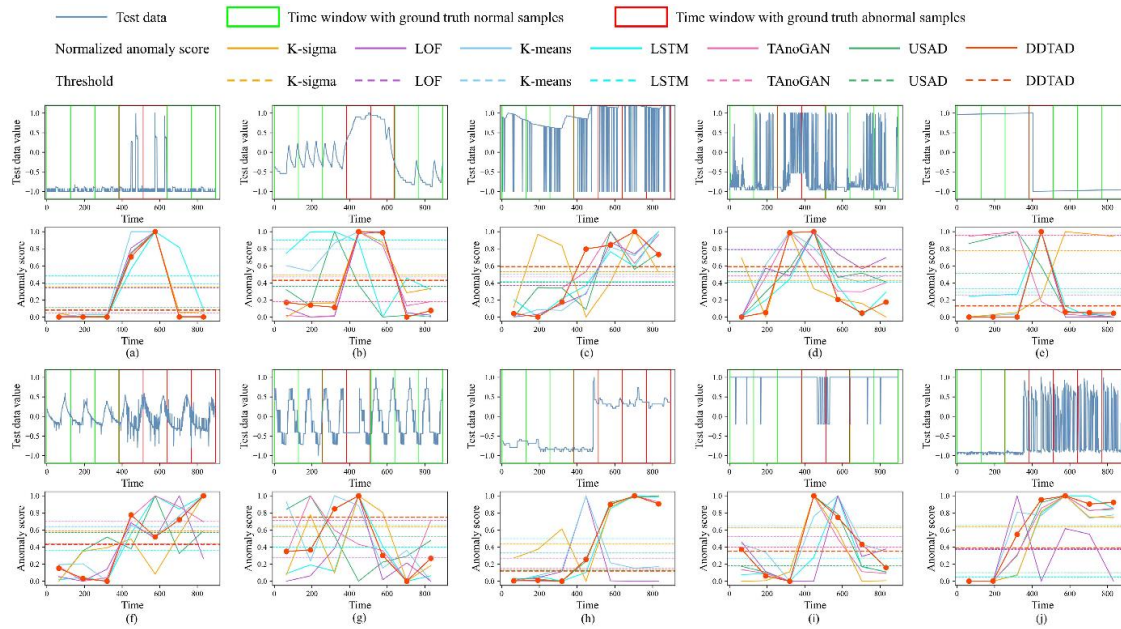


Figure 1 OPS-SAT CubeSat [3]

\*School of Aerospace, Transport and Management, Cranfield University: Cranfield, England. **Corresponding Author:** sakshivittal01@gmail.com.  
**Article History:** Received: 02-Sep-2025 || Revised: 15-Dec-2025 || Accepted: 30-Dec-2025 || Published Online: 30-Dec-2025.

Typically operating in Low Earth Orbit (LEO), CubeSats house critical subsystems such as the Electrical Power System (EPS), Onboard Computer (OBC), communications, thermal control, and Attitude Determination and Control System (ADCS). These subsystems continuously generate telemetry data, which is transmitted to ground stations to monitor system performance and ensure mission success. Satellite telemetry data is essential for monitoring the operational status of various subsystems, diagnosing faults, and making command decisions. However, as the number of CubeSats in orbit increases, the volume and complexity of telemetry data have also grown. Manual inspection or rule-based threshold alerts are no longer sufficient for early detection of anomalies, especially under limited bandwidth and constrained contact windows [4].



**Figure 2 Example of an Anomalous Segment in a Telemetry Time-Series, Representative of the Real-World Data in OPSSAT-AD [4]**

Furthermore, CubeSats are prone to failures due to their small form factor and cost-driven design limitations. Common issues include battery degradation, thermal control failure, and hardware anomalies in attitude or power systems. In many cases, faults are preceded by subtle deviations in telemetry that go undetected without intelligent monitoring systems. Traditional rule-based monitoring systems are poorly suited for handling the scale and complexity of modern telemetry streams, especially in time-series and multi-channel cases. Consequently, machine learning (ML) has been proposed as a promising approach to enhance anomaly detection by learning from past telemetry data and generalizing to novel behaviors [4], [5]. However, a significant barrier to the adoption and validation of these ML approaches has been the scarcity of publicly available and well-annotated satellite telemetry datasets [6].

The publication of OPSSAT-AD [4] and ESA-ADB [6] addressed this gap by providing the first publicly accessible, real-world telemetry datasets with annotated anomalies. OPSSAT-AD, derived from the OPS-SAT CubeSat mission of the European Space Agency, contains 2,123 annotated fragments of time-series data from nine telemetry channels and serves as a benchmark for the evaluation of both traditional and deep learning methods. ESA-ADB builds upon this effort by aggregating telemetry from three ESA missions using hierarchical tagging and a structured time-series format. These datasets enable fair comparison between anomaly detection methodologies and facilitate reproducibility in experimental validation.



**Figure 3 Machine Learning Approach in CubeSat [Courtesy: AI, Gemini Pro]**

Numerous recent studies have employed these datasets to explore the performance of machine learning models in spacecraft anomaly detection. For instance, [Ruszcak et al. \[4\]](#) demonstrate that both classical models (e.g., logistic regression, random forest) and deep neural networks (e.g., temporal convolutional networks, LSTMs) can achieve high performance on OPS-SAT telemetry, with some configurations exceeding 98% accuracy on unseen data. This contrasts with more general reviews, such as that by [Fejjari et al. \[7\]](#), which place these findings within the broader framework of spacecraft anomaly detection and discuss open challenges, including real-time deployment, interpretability, and robustness to noise. This case study investigates the suitability of recent ML approaches for CubeSat telemetry anomaly detection by analyzing their performance and underlying assumptions in the context of real-world datasets. Specifically, we examine the methodological foundations and evaluation strategies employed in literature, assess the generalization capabilities of leading models, and highlight directions for future research, particularly in on-board implementation and dataset standardization.

## **2. Traditional Vs. Machine Learning (ML) Based Anomaly Detection Approaches**

Traditionally, CubeSat anomaly detection has relied on rule-based or fixed-threshold methods derived from engineering models. For example, a warning is generated when the battery voltage drops below 6.8 V. While these methods are interpretable and uncomplicated, they lack flexibility and cannot identify multivariate associations or time-varying correlations. In addition, hardcoded rules are susceptible to environmental changes (e.g., eclipse phases) or component aging. On the other hand, ML-based methods offer a data-driven approach to anomaly detection by learning normal behavioral patterns from historical telemetry data. ML models can detect subtle or previously unknown anomalies without requiring explicit definitions, which is a significant advantage for autonomous or long-duration missions. These methods fall into three classes:

- **Supervised learning**, using labelled examples of both normal and anomalous data.
- **Unsupervised learning**, where only normal data is assumed and anomalies are identified as outliers.
- **Semi-supervised learning**, in which the model is primarily trained on normal data but can leverage a limited amount of labelled information.

ML algorithms generally offer improved generalization, can handle high-dimensional inputs, and can adapt to system changes. However, they require careful tuning, large volumes of data, and rigorous validation to avoid false alarms or missed detections.

### 3. Overview of Relevant ML Models

Several ML models have been applied to spacecraft telemetry anomaly detection, each with strengths and trade-offs:

#### i) One-Class Support Vector Machine (OC-SVM):

A classical unsupervised algorithm that defines a decision boundary around normal data. It is fast and interpretable but struggles with high-dimensional or temporal data [8].

#### ii) Isolation Forest:

An ensemble method that detects anomalies based on how easily they can be isolated in random partitions. Effective for tabular data but lacks temporal awareness [7].

#### iii) Autoencoders (AE):

Neural networks trained to reconstruct input telemetry. Anomalies are detected when reconstruction error exceeds a threshold. Suitable for both univariate and multivariate data [9].

#### iv) Long Short-Term Memory (LSTM):

A recurrent neural network (RNN) variant that models long-term dependencies in time-series data. Often used to predict next telemetry values or reconstruct sequences [10].

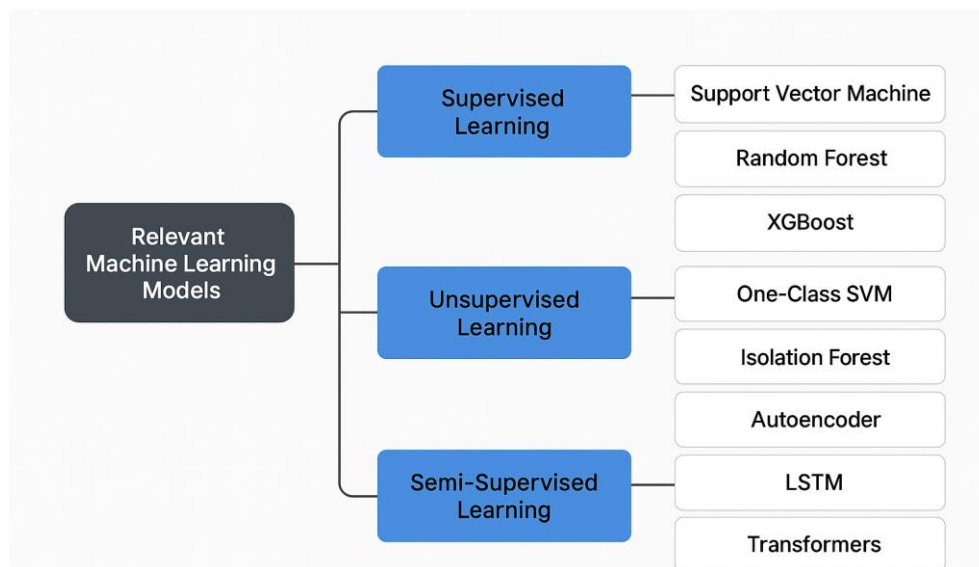
#### v) Temporal Convolutional Networks (TCN):

1D convolutional models that capture both short and long-range dependencies. Faster training than RNNs and easier to parallelise [11].

#### vi) Transformer Models and Graph Neural Networks (GNN):

Emerging models that learn complex relationships across telemetry channels (e.g., inter-subsystem dependencies). These models show strong results but require substantial data and computation [6] [12].

Each model type is evaluated using benchmark datasets such as OPSSAT-AD and ESA-ADB, where labelled anomalies and standard evaluation metrics (F1-score, precision, recall) support comparative analysis.



**Figure 4 Classification of Machine Learning Models**



#### 4. Datasets

Recent advancements in machine learning-based anomaly detection for satellite telemetry have been significantly bolstered by the release of two high-quality, publicly available datasets: OPSSAT-AD and ESA-ADB. These datasets are among the first to provide labelled telemetry data from real satellite missions, enabling standardised evaluation and reproducibility of ML algorithms in the space domain.

##### OPSSAT-AD Dataset:

The OPSSAT-AD (Anomaly Detection) dataset was derived from the ESA OPS-SAT CubeSat, a 3U mission dedicated to in-orbit experimentation with AI technologies. Released in 2024, this dataset contains 2,123 univariate time-series fragments sampled from nine distinct telemetry channels, such as battery voltage, solar current, gyroscope readings, and temperature sensors [5]. Approximately 20% of the fragments are labelled as anomalous, based on expert annotations and operational event logs.

Each fragment in the dataset represents a fixed-length sequence of sensor values, pre-processed to remove outliers and clipped to operationally relevant ranges. The dataset also includes:

- **Labels:** binary labels indicating whether the sequence contains an anomaly.
- **Channel metadata:** descriptions of the monitored subsystems.
- **Feature extraction scripts:** accompanying notebooks demonstrate how to transform the raw time-series into features suitable for ML models.

The total dataset size is modest (~15 MB) due to its univariate and fragment-based structure, making it ideal for prototyping and benchmarking traditional and deep learning models. Baseline performance benchmarks using logistic regression, random forests, LSTM, TCN, and CNN models have been reported, with F1-scores exceeding 0.95 in the best configurations [4] [5].

##### ESA-ADB Benchmark:

The ESA Anomaly Detection Benchmark (ESA-ADB), released in 2024, is a significantly larger and more comprehensive dataset curated from multiple ESA spacecraft missions. It includes over 30 GB of telemetry time-series data across hundreds of channels from three different missions, two of which were used in the benchmarking study [6].

Unlike OPSSAT-AD, ESA-ADB provides multivariate telemetry with a high sampling frequency and manual anomaly annotations performed by spacecraft operations engineers. Each anomaly is annotated with start and end timestamps, a description, and a severity rating. This level of detail allows for realistic benchmarking of anomaly detection algorithms in time-localized, multichannel contexts.

##### Additional features include:

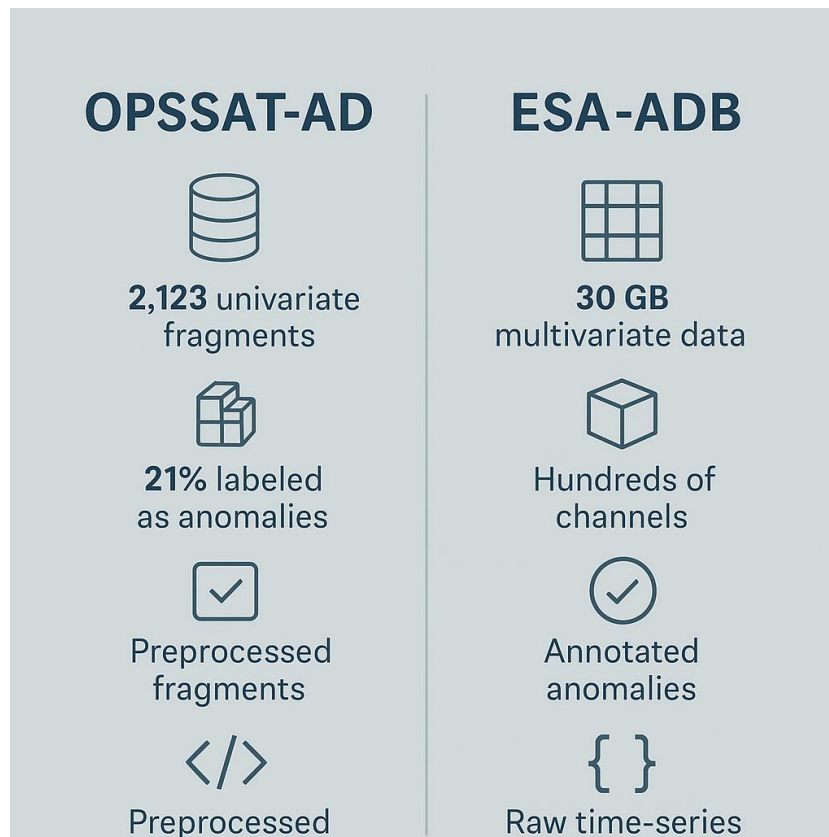
- **Multi-mission architecture** supports generalization studies across spacecraft.
- **TimeEval integration:** benchmarking suite with over 30 pre-implemented ML and statistical models.
- **Task configuration:** supports both online and offline anomaly detection scenarios.

Despite its richness, ESA-ADB poses several challenges for practitioners:

- **High dimensionality:** telemetry streams may include hundreds of correlated variables.
- **Label imbalance:** anomalies are rare (~1–5% of data), requiring robust sampling strategies.
- **Noise and missing data:** common in real-world telemetry, necessitating preprocessing such as interpolation or smoothing.
- **Computational demands:** due to its size and complexity, ESA-ADB typically requires GPU acceleration and optimized data pipelines for effective use.

Both OPSSAT-AD and ESA-ADB address a longstanding gap in public access to labelled satellite telemetry data, facilitating the development and evaluation of anomaly detection algorithms. OPSSAT-AD is ideal for quick prototyping on compact datasets, while ESA-ADB offers the scale and realism needed for industrial-strength model validation.





**Figure 5 OPSSAT-AD vs ESA-ADB [Courtesy: AI, Gemini Pro]**

## 5. Review of Existing Studies

Recent studies have evaluated a range of traditional, ensemble, and deep learning models, reporting varying degrees of success based on evaluation metrics such as accuracy, F1-score, precision, recall, and area under the ROC curve (AUC).

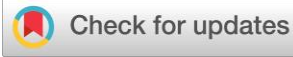
### Performance of Key Models:

Ruszczak et al. (2024) [4] used the OPSSAT-AD dataset to benchmark over 30 machine learning algorithms, including logistic regression, decision trees, random forest, support vector machines, k-nearest neighbours, and deep models like convolutional neural networks (CNNs), long short-term memory (LSTM), and temporal convolutional networks (TCN). Key performance results from their study include:

- Random Forest achieved an F1-score of 0.987, showing strong robustness across multiple telemetry channels.
- Logistic Regression demonstrated near-baseline performance with an F1-score of 0.943, despite its simplicity.
- CNN and TCN models achieved F1-scores > 0.95, suggesting that deep models can effectively capture temporal features even in univariate telemetry fragments.
- One-Class SVM underperformed (F1-score < 0.80), reflecting limitations in handling noisy telemetry.

Similarly, Kotowski et al. (2024) [6] introduced the ESA-ADB benchmark and evaluated several multivariate anomaly detection methods using the TimeEval benchmarking suite. They reported:

- LSTM models trained in a reconstruction setting performed well for long-horizon temporal modelling, achieving AUC scores above 0.90 on several channels.
- Isolation Forest achieved moderate F1-scores (~0.72), indicating its effectiveness when anomaly patterns are clearly distinguishable.
- Autoencoders (AE) showed variable performance depending on channel correlation and the presence of noise.
- Graph-based models (e.g., GNNs) showed promise in capturing cross-channel dependencies but suffered from long training times and interpretability issues.



Fejjari et al. (2025) provided a comprehensive meta-analysis of ML-based anomaly detection methods in spacecraft telemetry [7]. Their synthesis of performance results from over 20 papers indicates:

- Supervised methods generally outperform unsupervised methods when labelled data is available but are susceptible to overfitting due to class imbalance.
- Unsupervised methods are robust to new mission data but may yield higher false positive rates, especially on noisy telemetry.

Deep learning methods like LSTM and TCN consistently outperform shallow models in multivariate and temporal contexts, though they require larger training data and tuning.

#### Advantages and Disadvantages of Approaches:

**Table 1 Advantages and Disadvantages of Machine Learning Models**

Model Type	Advantages	Disadvantages
<b>Random Forest</b>	High accuracy, easy to train, handles non-linear patterns	Less effective on temporal data
<b>Logistic Regression</b>	Simple, interpretable	Poor temporal modelling
<b>LSTM</b>	Captures long-term dependencies, good for sequence modelling	Requires tuning, longer training, sensitive to noise
<b>TCN</b>	Fast, parallelizable, good for time-series	Needs fixed input size, limited contextual memory
<b>Autoencoder</b>	Useful for reconstruction-based detection	Prone to false negatives with complex anomalies
<b>OC-SVM</b>	Simple unsupervised baseline	Poor scalability, weak in multivariate or noisy contexts
<b>GNN/Transformer</b>	Captures inter-channel dependencies, scalable	High computational load, lower interpretability

#### Best Practices and Challenges Identified

From the reviewed literature, several best practices and recurring issues emerge:

##### Best Practices:

- **Windowed preprocessing:** Segmenting telemetry into fixed-size fragments improves detection performance across models [4] [6].
- **Channel-wise modelling:** Training separate models per telemetry channel reduces false positives and enhances interpretability [4].
- **Feature standardisation:** Scaling and normalization are essential to reduce model bias and improve convergence [7].

##### Challenges:

- **Overfitting:** Especially in supervised models trained on small or unbalanced datasets (common in OPSSAT-AD).
- **Lack of generalisation:** Models trained on one CubeSat or mission often fail to generalize to another without retraining [7].
- **Anomaly sparsity:** Both datasets are imbalanced, with anomalies constituting less than 5-20% of data, requiring resampling or weighted loss functions [6].
- **Interpretability:** Deep models, while accurate, offer limited explainability-prompting interest in explainable ML methods [12].

---

## 6. Discussion

### Strengths and Limitations of Current Datasets:

The introduction of OPSSAT-AD and ESA-ADB has marked a pivotal shift in the field of CubeSat telemetry anomaly detection by providing real-world, labeled telemetry datasets suitable for machine learning experimentation. These datasets offer several strengths:

- **Operational authenticity:** Both datasets originate from actual ESA missions, preserving realistic telemetry characteristics such as noise, sampling irregularity, and system complexity.
- **Standardised benchmarking:** By publishing baseline results and evaluation pipelines, these datasets support reproducibility and allow fair comparison of different models.
- **Annotation richness:** ESA-ADB, in particular, provides timestamped anomaly intervals and descriptions curated by mission engineers, enabling fine-grained analysis.

However, important limitations remain:

- **Data sparsity and imbalance:** Both datasets exhibit low anomaly-to-normal ratios, posing challenges for supervised models and increasing the risk of overfitting.
- **Fragmentation:** OPSSAT-AD, while clean and compact, lacks full-mission continuity and may not reflect long-range dependencies seen in live operations.
- **Mission-specific bias:** Models trained solely on these datasets may fail to generalize to telemetry from different missions, sensors, or operational regimes.

### Realism of Benchmark Tasks vs Operational Needs:

While the datasets are derived from operational spacecraft, the benchmark tasks (e.g., detecting anomalies in pre-processed fragments) may not fully simulate the online, real-time constraints encountered in actual mission control environments. In practice, operators must:

- Detect anomalies in streaming telemetry, often with limited prior context.
- Prioritize detection latency and explainability over raw accuracy.
- Integrate ML outputs with existing fault management systems and human-in-the-loop decision processes.

Therefore, while OPSSAT-AD and ESA-ADB are invaluable for offline evaluation, there remains a disconnect between current benchmarks and the real-world deployment environment.

### Most Effective ML Models and Underlying Rationale:

Based on the reviewed studies, several observations can be made regarding model performance:

- TCNs and LSTMs consistently outperform shallow models (e.g., OC-SVM, Isolation Forest) on time-dependent data due to their ability to capture temporal dependencies [4] [6].
- Random Forests and XGBoost remain strong contenders for univariate or short-window tasks, offering robustness and low latency [4].
- Autoencoders perform well in reconstruction-based settings, especially when paired with denoising or attention mechanisms.
- Graph Neural Networks and Transformer-based models have shown promise for modelling cross-channel dependencies, though they remain computationally intensive and harder to interpret [12].

The effectiveness of each model often correlates with:

- Data volume and structure (e.g., multivariate vs univariate)
- Availability of labels
- Deployment requirements (onboard vs ground-based inference)

### Future Research Directions:

Despite recent progress, several promising research directions remain underexplored:

- **Transfer Learning Across Missions:** To improve generalisability, future models should leverage transfer learning techniques that can adapt from one satellite or subsystem to another with minimal retraining. Pretraining on ESA-ADB and fine-tuning on mission-specific data is a potential strategy.
-





- **Onboard and Edge Deployment:** Lightweight anomaly detection models-particularly those based on decision trees or quantized neural networks-could be optimized for real-time, onboard inference under hardware constraints typical of CubeSats.
- **Explainability and Human-in-the-Loop Systems:** Developing interpretable ML models is critical to gaining operator trust and facilitating operational integration. Techniques such as SHAP, attention visualisation, and anomaly attribution offer a pathway toward explainable anomaly detection [7].
- **Hybrid and Ensemble Methods:** Combining supervised and unsupervised models, or integrating statistical with ML-based detectors, could enhance both precision and recall, especially in edge-case scenarios.
- **Synthetic Data Augmentation:** Simulated anomaly injection, GAN-based telemetry synthesis, and domain adaptation could address data scarcity and imbalance challenges.

## 7. Conclusion

This case study explored the current landscape of anomaly detection in CubeSat telemetry using machine learning, with a particular focus on recent advancements enabled by the OPSSAT-AD and ESA-ADB benchmark datasets. These publicly available datasets represent a significant step forward for the space systems community by providing real-world, labelled telemetry suitable for reproducible evaluation of anomaly detection algorithms. Our review of existing studies highlights that deep learning models particularly LSTMs and temporal convolutional networks are highly effective at capturing temporal and multivariate dependencies in telemetry streams. At the same time, traditional models such as Random Forests and autoencoders remain competitive in constrained or univariate settings. Performance metrics reported in the literature indicate that several ML approaches now achieve F1-scores above 0.95 on structured telemetry fragments, suggesting that ML-based anomaly detection is reaching operational maturity in offline contexts.

Despite this progress, significant challenges remain. These include the limited generalisability of models across missions, the imbalance and sparsity of anomalies in real telemetry, and the lack of interpretability in complex deep learning systems. Moreover, a noticeable gap persists between benchmark-based model validation and the requirements of real-time, onboard anomaly detection in operational environments. This case study contributes to the CubeSat anomaly detection community by synthesising recent research findings, evaluating the effectiveness of available datasets, and identifying key trade-offs in model selection and deployment. It also outlines several promising future directions such as transfer learning, onboard inference, and explainable AI that could help bridge the gap between research and operational integration. Ultimately, the future of ML-based anomaly detection in small satellite systems depends on sustained efforts to expand and standardise benchmark datasets, incorporate operational constraints into model design, and validate algorithms in real flight scenarios. Collaboration among satellite operators, data scientists, and ML researchers will be essential to ensure that these technologies deliver robust and trustworthy performance in mission-critical environments.

## 8. References

- [1] Puig-Suari, J., Turner, C., & Ahlgren, W. (2001). Development of the standard CubeSat deployer and a CubeSat class PicoSatellite. In 2001 IEEE Aerospace Conference Proceedings (Cat. No.01TH8542) (Vol. 1, pp. 1-347). IEEE. <https://doi.org/10.1109/AERO.2001.985275>
- [2] Cappelletti, C., Battistini, S., & Malphrus, B. K. (Eds.). (2020). Cubesat handbook: From mission design to operations. Academic Press.
- [3] European Space Agency. (n.d.). Accessed from <https://www.esa.int/>.
- [4] Ruszczak, B., Kotowski, K., Evans, D., & Nalepa, J. (2025). The OPS-SAT benchmark for detecting anomalies in satellite telemetry. *Scientific Data*, 12(1), Article 710. <https://doi.org/10.1038/s41597-025-05035-3>.
- [5] Cuéllar, S., Santos, M., Alonso, F., Fabregas, E., & Farias, G. (2024). Explainable anomaly detection in spacecraft telemetry. *Engineering Applications of Artificial Intelligence*, 133, Article 108083. <https://doi.org/10.1016/j.engappai.2024.108083>.
- [6] Kotowski, K., Haskamp, C., Andrzejewski, J., Ruszczak, B., Nalepa, J., Lakey, D., ... & De Canio, G. (2024). European space agency benchmark for anomaly detection in satellite telemetry. <https://doi.org/10.48550/arXiv.2406.17826>.
- [7] Fejjari, A., Delavault, A., Camilleri, R., & Valentino, G. (2025). A review of anomaly detection in spacecraft telemetry data. *Applied Sciences*, 15(10), Article 5653. <https://doi.org/10.3390/app15105653>.
- [8] Cespedes, A. J. J., Pangestu, B. H. B., Hanazawa, A., & Cho, M. (2022). Performance evaluation of machine learning methods for anomaly detection in CubeSat solar panels. *Applied Sciences*, 12(17), Article 8634. <https://doi.org/10.3390/app12178634>.
- [9] He, J., Cheng, Z., & Guo, B. (2022). Anomaly detection in satellite telemetry data using a sparse feature-based method. *Sensors*, 22(17), Article 6358. <https://doi.org/10.3390/s22176358>.
- [10] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 387-395). <https://doi.org/10.1145/3219819.3219845>.
- [11] Liu, L., Tian, L., Kang, Z., & Wan, T. (2023). Spacecraft anomaly detection with attention temporal convolution networks. *Neural Computing and Applications*, 35(13), 9753-9761. <https://doi.org/10.1007/s00521-023-08213-9>.
- [12] Cuéllar, S., Santos, M., Alonso, F., Fabregas, E., & Farias, G. (2024). Explainable anomaly detection in spacecraft telemetry. *Engineering Applications of Artificial Intelligence*, 133, Article 108083. <https://doi.org/10.1016/j.engappai.2024.108083>.

## 9. Conflict of Interest

The author declares no competing conflict of interest.

## 10. Funding

No funding was issued for this research.