74th International Astronautical Congress (IAC), Baku, Azerbaijan, 2-6 October 2023.

IAC-23,B6,3,5,x78668

Look ma, no ground truth! On building supervised anomaly detection from OPS-SAT telemetry

Jakub Nalepa[ab]*, Bogdan Ruszczak[ac], Krzysztof Kotowski[a], Jacek Andrzejewski[a], , Alicja Musial[a], David Evans[d], Vladimir Zelenevskiy[de], Sam Bammens[d], Rodrigo Laurinovics[d]

[a] *KP Labs, Gliwice, Poland*, {jnalepa, bruszczak, kkotowski}@kplabs.pl
[b] *Department of Algorithmics and Software, Silesian University of Technology, Gliwice, Poland*
[c] *Department of Informatics, Opole University of Technology, Opole, Poland*
[d] *European Space Agency/ESOC, Darmstadt, Germany*
[e] *Telespazio Germany GmbH, Darmstadt, Germany*
[f] *IrbGS ltd, "Viraki", Ance Parish, Irbene, Latvia*
* Corresponding Author

Abstract

Detecting anomalies in satellite telemetry data is pivotal in ensuring its safe operations. Although there exist various data-driven techniques for determining abnormal parts of the signal, they are virtually never validated over real telemetries. Analyzing such data is challenging due to its intrinsic characteristics, as telemetry may be noisy and affected by incorrect acquisition, e.g., rendering missing parts of the signal. Although there exist data-driven algorithms for detecting abnormal events, spanning across classic techniques exploiting expert systems, unsupervised approaches and deep learning models, they are virtually never validated on real-life telemetry data. Also, they often require long time-series data to build a model that reflects the nominal operation of the satellite – capturing it on board is tedious and time-consuming, thus data-level digital twins have been blooming to simulate the correct telemetry. Benchmark datasets commonly exploited to validate detection algorithms contain time-series data, where each time series is split into its training and test parts, presenting similar characteristics. Such data is not affected by the practical challenges commonly observed in on-board telemetry, such as data noisiness or missing data, due to e.g., inappropriate signal acquisition. Therefore, the estimated anomaly detection capabilities of data-driven techniques may easily become over-optimistic, and the experimental scenarios are often flawed by methodological issues in the field. In this paper, we present our approach toward building a supervised machine learning model for detecting anomalies in real-life OPS-SAT telemetry data. We will discuss our procedure for building a labeled dataset of telemetry examples from a large dataset of unlabeled telemetry, and will present the importance of following a rigorous procedure for this task, as the quality of the ground-truth annotations (elaborated by a data scientist, with and without the label validation performed by the OPS-SAT Operations Team) affects not only the training process, but also the final validation of the machine learning model. We will thoroughly discuss our quantitative and qualitative experimental analysis that allowed us to objectively quantify the detection capabilities of the models, benefiting from hand-crafted feature extractors and classical supervised learners, even for (extremely limited) ground truth data. Finally, we will discuss the approach that we followed to prepare a resulting machine learning model for deploying it on-board OPS-SAT.

Keywords: machine learning, anomaly detection, telemetries, timeseries, on-board processing, data augmentation.

1. Introduction

Detecting irregularities within the telemetry data collected from satellites is a critical step in ensuring their safe operation and facilitating rapid response to various malfunctions and threats [1–5, 9, 12]. Generally, three primary types of anomalies need consideration in complex missions:

- **Point Anomalies**: These occur when telemetry values fall outside the expected operational range.
- **Collective Anomalies**: These involve sequences of consecutive telemetry values that collectively exhibit anomalies, with a single data point not necessarily considered an anomaly.
- **Contextual Anomalies**: Single telemetry values exhibit anomalies within their local context.

Creating detection systems for such unexpected events demands substantial expert knowledge to define the typical ranges for specific telemetry channels. These systems are often custom-designed for a particular mission and may not readily apply to future spacecraft.

To automate the telemetry data anomaly detection process, several research avenues have been explored in the literature. "Out-of-limit" detection engines verify if actual telemetry values fall within predefined ranges and are effective in identifying point anomalies. Expert systems have also been proposed for handling other types of events [5]. While machine learning algorithms

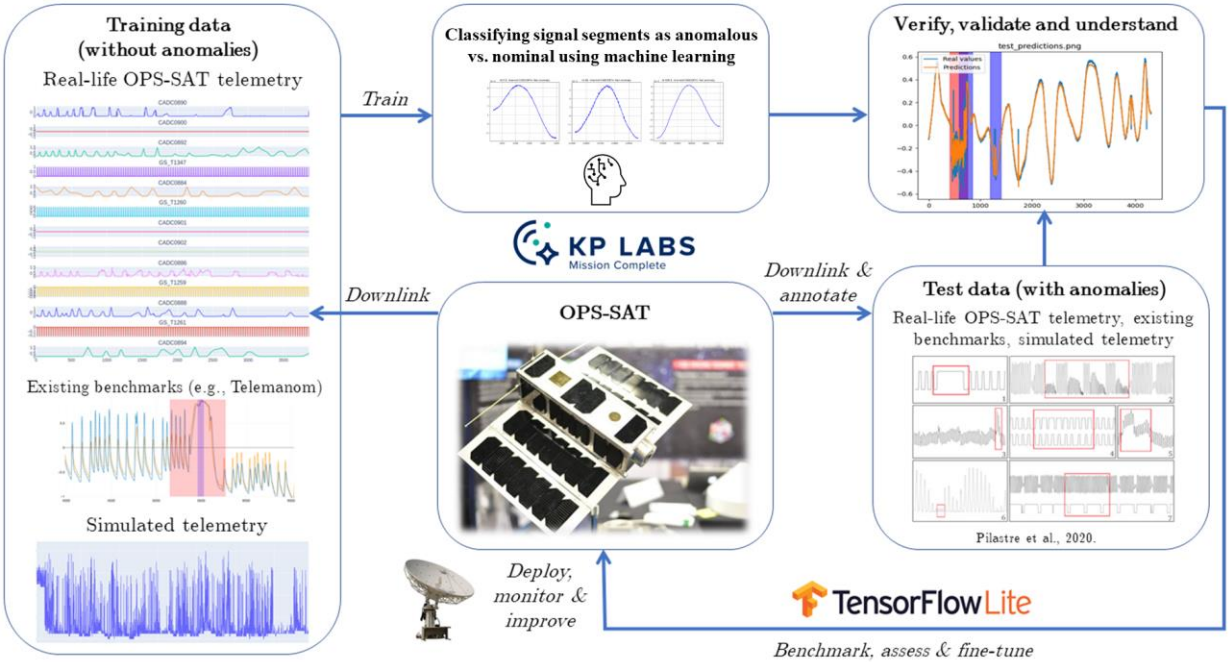

Fig. 1. A high-level flowchart showing how a machine learning model for anomaly detection from satellite telemetry could be deployed for the real-world data.

have been developed for telemetry data anomaly detection (covering both supervised and unsupervised techniques [6–9]), they tend to be highly parameterized, and improperly tuned hyperparameters can significantly impact their performance ([10, 11]).

In recent years, deep learning has achieved remarkable success in various scientific and industrial domains, including the detection of anomalies in sequential data. Artificial neural networks known as recurrent neural networks have shown promise in forward prediction based on time series inputs. RNN architectures based on Long Short-Term Memory or Gated Recurrent Unit modules are particularly well-suited for this purpose due to their feedback connections and temporal learning capabilities, which enable them to refine predicted signals and compare them with actual telemetry data. By analyzing the differences between these signals, we can identify anomalous events.

In this paper, we focus on leveraging recent advances in the field and employ machine learning anomaly detection techniques to detect anomalies in real OPS-SAT telemetry signals. Applying such data-driven approaches to real-world data poses, however, several important challenges, with the most important one concerned with the availability of the ground-truth data (i.e., telemetry signals accompanied with the ground-truth information about the corresponding anomalous events). Here, we present our approach toward developing a dataset – in cooperation with the OPS-SAT Operations team – that could be used not only to build machine learning models for detecting anomalous events from telemetry data, but also for thoroughly verifying such techniques over real-world data.

The remainder of this paper is structured as follows. In Section 2, we present our approach toward anomaly detection in the satellite telemetry data, and discuss our dataset, which is used to experimentally verify the operational capabilities of the presented approach. Then, in Section 3, we elaborate on the obtained experimental results. Finally, Section 4 concludes the paper and shows the research directions that could emerge from the research presented in this paper.

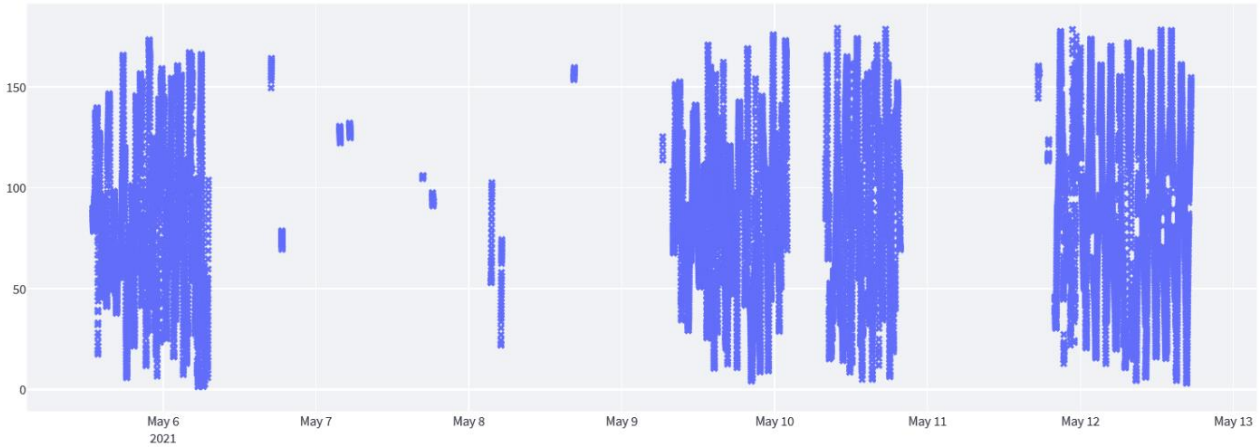

Fig. 2. Example of a telemetry signal with hours-long readout gaps.

Fig. 3. The idea of telemetry segments (annotated with different colors).

2. Materials and methods

In this section, we discuss our real-world dataset that was used to build and verify the machine learning algorithms for detecting anomalous events in the OPS-SAT telemetry (Section 2.1). In Section 2.2, we discuss the supervised learners utilized for this task.

2.1 Building OPS-SAT telemetry dataset for anomaly detection

In this research, we leverage a dataset of telemetry signals derived from real-life OPS-SAT telemetry data. The selection of signals for this investigation was guided by input from OPS-SAT operators, who highlighted telemetry parameters of practical importance. Consequently, we considered several telemetry readings from the magnetometer and the PD channels. The resulting dataset initially contained 524 identified anomalies. However, after a careful dataset review, we retained 445 anomalies of various types. To annotate and analyze the telemetry signal data, we employed the KP Labs' OXI tool for telemetry analysis, which can be accessed at https://oxi.kplabs.pl/ [12].

An initial examination of the downloaded telemetry signals unveiled certain challenges that informed our choice of the target algorithm for implementation. These challenges included data gaps, discontinuities within the signal, abnormal signal patterns (which are not necessarily anomalous events that should be detected by the automated anomaly detection algorithm), significant changes in signal characteristics (which may be considered a "concept drift"), noisy sections of specific telemetry channels, periodic signals, and segments devoid of apparent periodicity. An example illustrating noisy and fragmented telemetry is provided in Fig. 2. In light of these real-world complexities, our dataset was meticulously curated by a committee comprising OPS-SAT operators and specialists in machine learning and data analysis. The final dataset was derived from the periodic segments of telemetry data spanning several months and encompassing various channels. These periodic signals were subsequently divided into segments, each of which was labeled as either anomalous or non-anomalous. The rationale behind this segmentation approach is elucidated in Fig. 3, which illustrates a real OPS-SAT periodic telemetry fragment divided into short segments, with each segment color-coded to enhance visualization.

In our analysis, we manually selected a total of 2134 segments from the identified periodic signal parts spanning from January 1, 2022, to August 1, 2022. These segments were sourced from various channels, and during our examination, we identified 524 anomalies. A detailed breakdown of this dataset is presented in Table 1, while non-anomalous telemetry segments are exemplified in Fig. 4, and identified anomalies are showcased in Fig. 5.

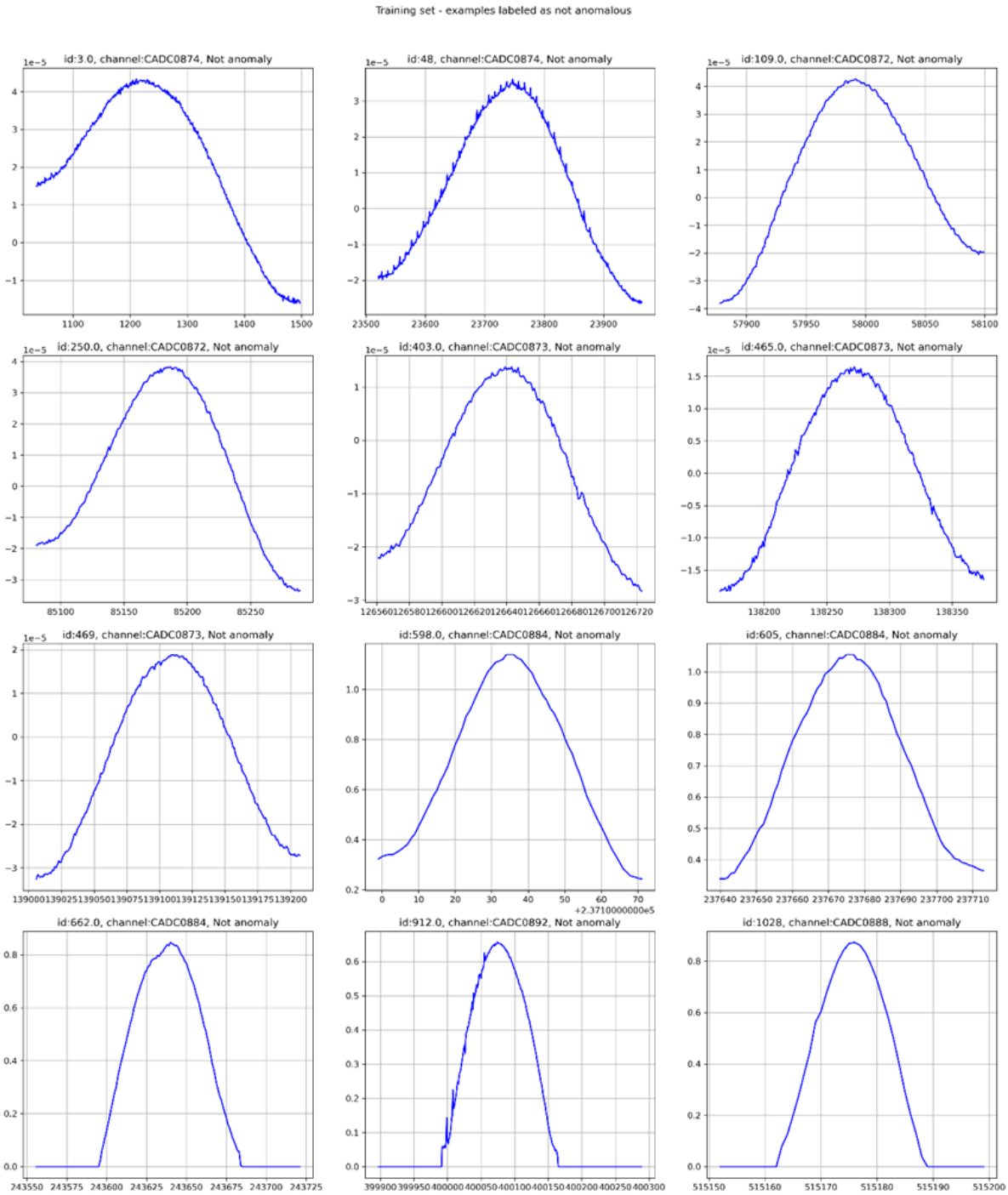

Fig. 4. Selected examples of non-anomalous segments of analyzed OPS-SAT telemetry channels.

Table 1. The preliminary training and validation sets (the number of anomalous and nominal segments) for anomalous and non-anomalous telemetry signals.

	Training	Validation	Total
Anomalies	396	128	**524**
Nominal	1205	405	**1610**
Total	**1601**	**533**	**2134**

Subsequently, we consulted with the OPS-SAT Operations team and refined the dataset to correct labels and ensure its accuracy. The updated dataset contains a reduced number of anomalies, as some segments were found to be non-anomalous upon further review. Table 2 provides a comprehensive summary of the updated dataset.

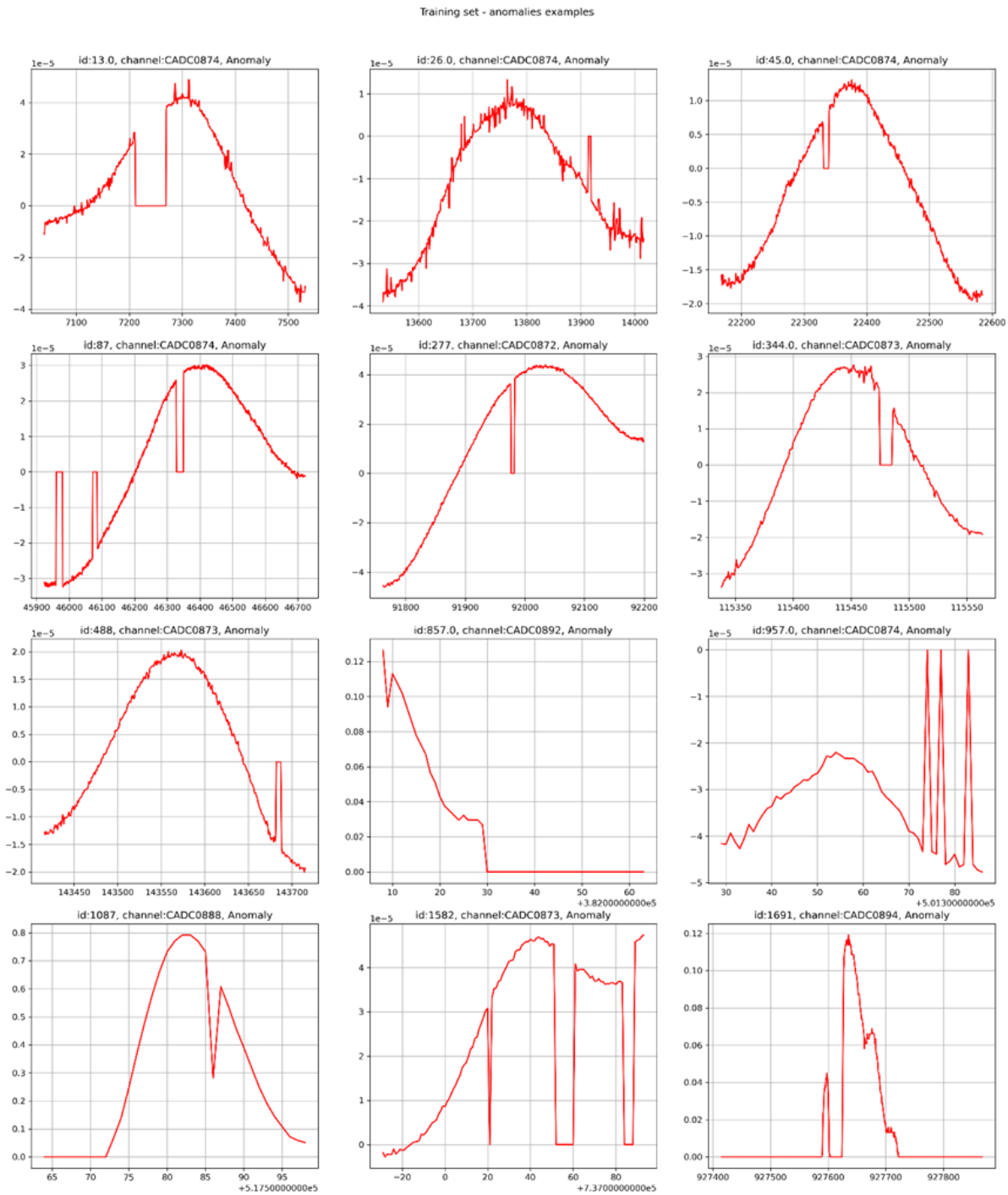

Fig. 5. Selected examples of anomalous segments of analyzed OPS-SAT telemetry channels.

To validate the performance of our anomaly detection algorithms, we prepared an additional test dataset. This dataset comprises telemetry fragments sourced from the same parameters as the training data but lacks periodicity. Table 3 provides an overview of this dataset, and Fig. 6 illustrates selected examples of labeled anomalous segments for reference.

Table 2. The updated training and validation sets for anomalous and non-anomalous telemetry segments.

	Training	Validation	Total
Anomalies	328	117	**445**
Nominal	1273	416	**1689**
Total	1601	533	**2134**

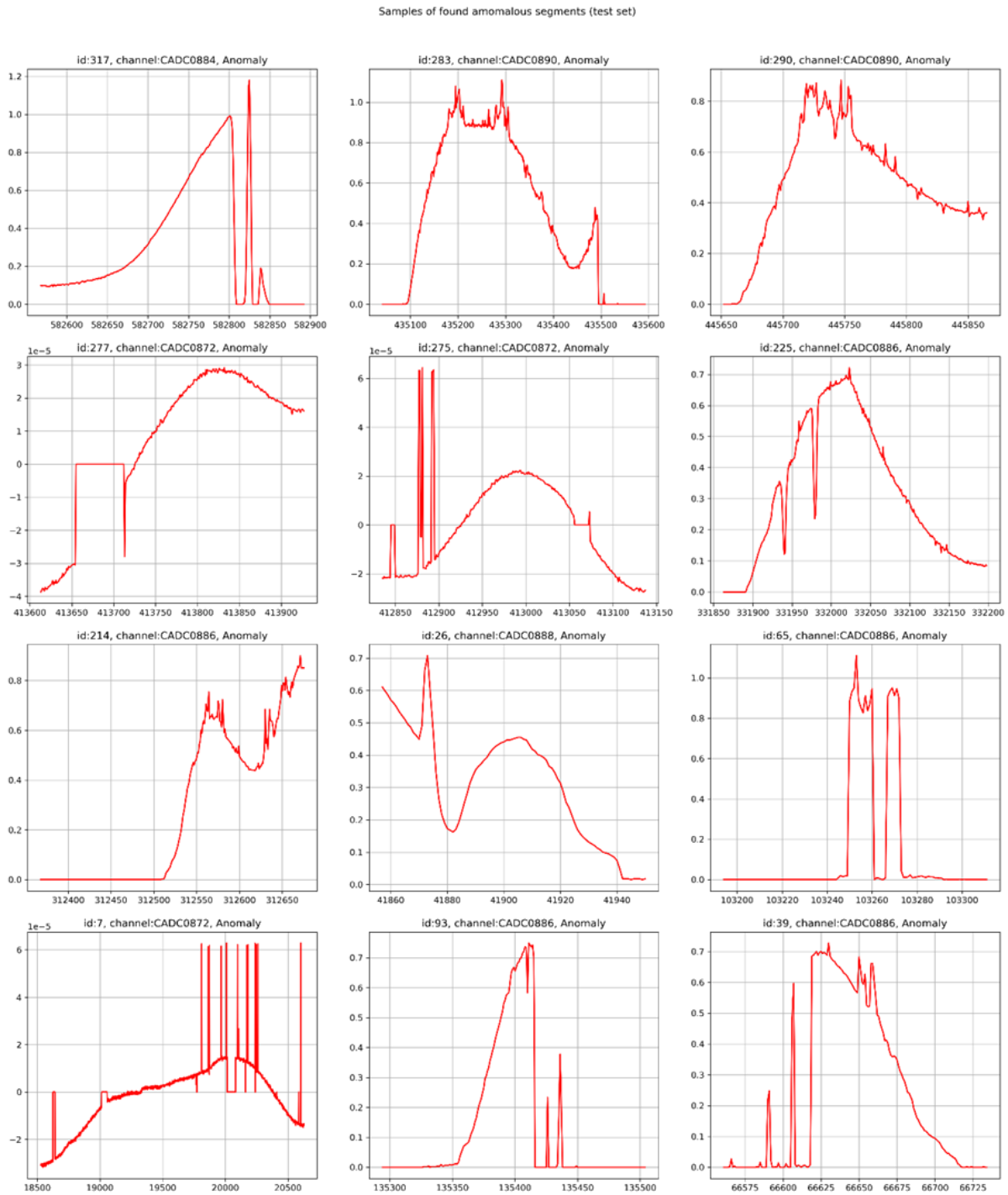

Fig. 6. Selected examples of anomalous segments of analyzed OPS-SAT telemetry channels.

Table 3. The test sets for anomalous and non-anomalous telemetry segments for OPS-SAT telemetries.

	Test
Anomalies	293
Nominal	50
Total	**343**

2.2 Detecting anomalies in OPS-SAT telemetry using supervised machine learning

For the purpose of anomaly detection in OPS-SAT telemetry signals, we employed a machine learning algorithm utilizing the random forest model architecture, as previously suggested in our prior

publication [13]. The random forest model was configured with the following settings:

- Maximum number of estimators: 25
- Quality criterion: Gini impurity
- Minimum number of samples required to split an internal node: 2
- Minimum number of samples required to be at a leaf node: 1

All features were considered for the best split, and there was no restriction on the maximum number of leaf nodes. No maximum number of samples for bootstrapping was imposed. In line with previous research [4, 5], which showed the effectiveness of ensemble machine learning models in classifying signals based on extracted features, we adopted the same approach. Here, we leveraged random forests for binary classification, wherein signals were categorized as either nominal or anomalous. The feature extraction process encompassed a range of characteristics, including the identification of signal peaks, detection of missing signal segments, and assessment of noise levels with the telemetry signal. Importantly, these feature extractors were designed to maintain robustness across telemetry segments of varying lengths, therefore can be conveniently exploited for telemetry signals of different lengths and characteristics.

We employed 18 distinct features for each segment, as suggested in [13]. These features span across basic segment statistics, including duration, length, mean, variance, and standard deviation. Additionally, we include the counts of peaks in the segment, both for the original segment and for slightly and strongly smoothed versions, peak counts for the first and second derivatives of the segment. Finally, we include the variance calculations for the first and second derivatives of the segment, the squared number of missing readouts, the weighted segment length relative to its sampling, the variance relative to the segment length and its duration.

3. Experimental results

In this section, we discuss the experimental results obtained over the validation set, without and with additional review of the dataset performed by the OPS-SAT Operations team (for the latter case, we additionally augment the training dataset using our data augmentation procedure [13]). To quantitatively analyze the capabilities of the algorithm, we exploited classical detection measures, including precision, recall, and the F1 score. Table 4 gathers the experimental results obtained for the validation dataset, showing that the supervised machine learning model can effectively deal with an appropriately curated yet limited in size training set for real-world anomaly detection from telemetry data.

Table 4. Preliminary results computed over the validation set (without additional review of telemetry segments performed by the OPS-SAT Operations team).

	Precision	Recall	F1 score
Anomalies	0.926	0.833	0.904
Nominal	0.964	0.978	0.971

In order to validate the accuracy of the labels generated by the machine learning team, we sought input from OPS-SAT operators. This validation process was conducted on a representative subset of segments. Given that the dataset primarily comprises non-anomalous segments, which are often straightforward to identify, our focus in this survey was on examining edge cases. For this validation exercise, we converted each segment into a discrete small image format. This approach facilitated a visual inspection by operators, enabling them to comprehensively review each segment, scrutinize finer details, and verify the labeling.

Table 5. The results computed using the validation set after the data augmentation and the dataset verification process.

	Precision	Recall	F1 score
Anomalies	0.929	0.897	0.913
Nominal	0.971	0.981	0.976

In the validation set, our model achieved an impressive accuracy of 0.957. When focusing specifically on anomaly classification, the precision and recall metrics for this task were 0.929 and 0.897, respectively. It's worth noting that the incorrectly classified segments in this set would also be challenging for human operators as well, given that they exhibit subtle characteristics that might suggest they are nominal. Conversely, the false positives in the classification include segments of the signal that could be interpreted as abnormal. Notably, the most difficult segments lie at the boundary between the anomalous and nominal parts of the telemetry channel.

In contrast to the training and validation datasets, which were constructed using periodic segments from the studied telemetry channels, the segments in the test set were sourced from other, non-periodic telemetries. This dataset was designed to assess the classifier's robustness in handling such signals. Remarkably, the classifier achieved an accuracy of 0.959 on this test set, along with precision and recall metrics of 0.968 and 0.959, respectively. These results indicate that the classifier can indeed cope with difficult telemetry signals, even when trained over relatively small training set with challenging characteristics.

4. Conclusions and outlook

In this paper, we investigated our end-to-end anomaly detection algorithm that was deployed for detecting anomalous events in the OPS-SAT telemetry data. Applying such techniques in practice is extremely challenging, due to the fact that the ground truth data does not exist before the satellite is operating in orbit. Here, we showed that the thorough analysis of the telemetry channels plays a crucial role in designing and developing datasets that could be utilized to build and verify machine learning models for the anomaly detection task. Our experimental results indicated that – albeit the extreme difficulty of the OPS-SAT telemetry data – the classic machine learning models can generalize over the unseen data and offer high-quality anomaly detection.

The outcomes of this project mark an exciting starting point for future research and development efforts in the field of telemetry data analysis. While we have made significant progress, several challenges remain, particularly at the telemetry data level. Some of these challenges include:

Under-representation of Specific Events: Addressing the under-representation of specific events in the data is crucial. Ensuring a more balanced dataset with adequate representation of all possible anomalies can improve the model's ability to detect rare events.

Extreme Imbalance in Training Datasets: Dealing with the extreme class imbalance often encountered in supervised learning scenarios is essential. Techniques like resampling, cost-sensitive learning, or synthetic data generation can be explored to mitigate this issue.

Few-Shot Learning: Leveraging recent advancements in few-shot learning could be beneficial. This approach enables models to generalize effectively even when trained on limited samples, making it useful for scenarios where collecting extensive labeled data is challenging.

FPGA Acceleration: Depending on the algorithm used, FPGA (Field-Programmable Gate Array) acceleration can significantly enhance on-board computational efficiency. FPGA hardware can be optimized to run specific machine learning algorithms, reducing computational load and power consumption.

By addressing these challenges and capitalizing on recent advancements in machine learning and hardware acceleration, we can strive to build more robust and efficient models for telemetry data analysis, further advancing the capabilities of satellite operations and anomaly detection.

Acknowledgements

This work was partially supported by the: “On-board Anomaly detection from the OPS-SAT telemetry using deep learning” project funded by the European Space Agency under contract No. 4000137339/22/NL/GLC/ov. JN was supported by the Silesian University of Technology grant for maintaining and developing research potential.

References

[1] U. Zahid et al., “FAT: Training Neural Networks for Reliable Inference Under Hardware Faults,” no. arXiv:2011.05873v1, 2020.

[2] X. Zhang, Y. Wang and W. Shi, “Performance Comparison of Machine Learning Packages on the Edges,” Boston, 2018.

[3] D. Velasco-Montero et al., “Performance Analysis of real-time DNN inference on Raspberry Pi,” vol. 10670, 2018.

[4] J. Nalepa, M. Antoniak, M. Myller, P. R. Lorenzo and M. Marcinkiewicz, “Towards Resource Frugal Deep Convolutional Neural Networks for Hyperspectral Image Segmentation,” Microprocess Microsyst, vol. 73, no. 102994, 2020.

[5] P. Benecki, S. Piechaczek, D. Kostrzewa, J. Nalepa, “Detecting Anomalies in Spacecraft Telemetry Using Evolutionary Thresholding and LSTMs”, Proc. GECCO 2021, pp. 143–144, ACM, 2021, https://doi.org/10.1145/3449726.3459411.

[6] K. Hundman et al., “Detecting Spacecraft Anomolies Using LSTMs and Nonparametric Dynamic Thresholding,” in KDD, London, 2018.

[7] Y. Gao et al., “An unsupervised Anomaly Detection Approach for Spacecraft Based on Normal Behaviour Clustering,” in ICICTA, 2012.

[8] K. Li et al. “A novel method for spacecraft electrical fault detection based on FCM clustering and WPSVM classification with PCA feature extraction.,” Proceedings of the Institution of Mechanical Engineers, J Aerosp Eng, pp. 98-108, 2017.

[9] S. Fuertes et al., “Improving spacecraft health monitoring with automatic anomaly detection techniques,” AIAA 2016-2430, GSCDP - Advanced Technologies for Space Operations V, 2016.

[10] C. Chang, “Satellite diagnostic system: An expert system for intelsat satellite operations,” 1992.

[11] J. Nalepa and M. Kawulok, “Selecting training sets for support vector machines: a review,” Artificial Intelligence Review, vol. 52, pp. 857-900, 2019.

[12] B. Ruszczak, K. Kotowski, J. Andrzejewski, C. Haskamp, J. Nalepa, “OXI: An online tool for visualization and annotation of satellite time series data”, SoftwareX, Volume 23, 2023, 101476, https://doi.org/10.1016/j.softx.2023.101476.

[13] B. Ruszczak et al., “Machine Learning Detects Anomalies in OPS-SAT Telemetry”. In: Mikyška et al., (eds) Computational Science – ICCS 2023. Lecture Notes in Computer Science, vol 14073. Springer. doi.org/10.1007/978-3-031-35995-8_21.