

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/378214516>


# Explainable anomaly detection in spacecraft telemetry

Article in Engineering Applications of Artificial Intelligence · February 2024  
 DOI: 10.1016/j.engappai.2024.108083

CITATIONS  
 17

READS  
 509

5 authors, including:




**Sara Cuéllar Carrillo**

Pontifical Catholic University of Valparaíso

10 PUBLICATIONS 63 CITATIONS

SEE PROFILE




**Matilde Santos Peñas**

Complutense University of Madrid

199 PUBLICATIONS 2,381 CITATIONS

SEE PROFILE




**Ernesto Fabregas**

National University of Distance Education

83 PUBLICATIONS 1,092 CITATIONS

SEE PROFILE



**Gonzalo Farias**

Pontifical Catholic University of Valparaíso

165 PUBLICATIONS 2,078 CITATIONS

SEE PROFILE



## Research paper

## Explainable anomaly detection in spacecraft telemetry

Sara Cuéllar<sup>a</sup>, Matilde Santos<sup>b</sup>, Fernando Alonso<sup>c</sup>, Ernesto Fabregas<sup>d</sup>, Gonzalo Farias<sup>a,\*</sup><sup>a</sup> Escuela de Ingeniería Eléctrica, Pontificia Universidad Católica de Valparaíso, Av. Brasil 2147, Valparaíso, 2362804, Chile<sup>b</sup> Instituto de Tecnología del Conocimiento, Universidad Complutense de Madrid, Profesor García Santesmases 9, Madrid, 28040, Spain<sup>c</sup> Head of Planning Systems Section, Telespazio Germany GmbH, Europaplatz 5, Darmstadt, D-64293, Germany<sup>d</sup> Departamento de Informática y Automática, Universidad Nacional de Educación a Distancia, c/Juan del Rosal, 16, Madrid, 28040, Spain

## ARTICLE INFO

## Keywords:

Anomaly detection

Spacecraft telemetry

Explainable machine learning

Time-series analysis

## ABSTRACT

As spacecraft missions become more complex and ambitious, it becomes increasingly important to track the status and health of the spacecraft in real-time to ensure mission success. Anomaly detection is a crucial part of spacecraft telemetry analysis, allowing engineers to quickly identify unexpected or abnormal behaviour reflected on spacecraft data and take appropriate corrective action. Traditional statistical methods based on threshold setting are often inadequate for detecting anomalies in this context, requiring the development of more sophisticated techniques that can handle the high-dimensional, non-linear, and non-stationary nature of spacecraft telemetry data such as machine learning-based techniques. This article presents an approach for anomaly detection using machine-learning techniques for spacecraft telemetry. The identification of anomaly types present on two real telemetry datasets from NASA is performed to incorporate information of magnitude, frequency, and waveform from known anomalies into the feature extraction process. Then, a machine-learning-based model is trained with the obtained features and tested with unknown real data. The proposed method achieves 95.3% of precision and 100% of Recall, giving a  $F_{0.5}$  score of 96.2% in both datasets, outperforming the metrics obtained on the existing related works, demonstrating that the inclusion of known anomalies can improve the performance of the data-driven models. Finally, an explainability analysis is performed to understand why a particular data instance has been identified as anomalous, proving the effectiveness of the feature extraction process.

## 1. Introduction

Spacecraft's complex structure together with a variable space environment, including sudden changes in temperature, direct radiation, and risk of collision with space debris, among others, impose a challenging control task. Improving the reliability of space system components is not enough to eliminate a possible failure. Additionally, the distance between the spacecraft and the ground station sometimes makes it impossible to inspect and repair a damaged component (Gao et al., 2012).

The data from the monitoring of these systems are therefore essential to ensure their proper functioning as far as possible to address any failures that are detected as soon as possible. The communication subsystem (C&DH) is responsible for acquiring measurements from the sensors on board the satellite and transmitting information about the status of all subsystems to the ground station. The analysis of this data is key to detecting and diagnosing any irregular behaviour that could cause the failure of a subsystem (Li et al., 2019).

Telemetry data is composed of unstable time series containing hundreds of sensor measurements representing the state of each spacecraft subsystem. Those time series are multivariate, heterogeneous, and multi-modal (Li et al., 2010).

Monitoring the status of satellites focuses on the early detection of anomalies, which is essential to avoid critical situations such as loss of control of the spacecraft or a serious failure. An anomaly is an instance that differs significantly from a predefined concept of normality (Ruff et al., 2021). Three types of anomalies can be found in the literature. Point anomalies are instances that are individually abnormal relative to other samples. Contextual anomalies are abnormal patterns that occur within a particular setting or environment but are not present under other circumstances; for example, a high measurement of solar radiation when the satellite is in an eclipse. Finally, collective or group anomalies are a set of instances that are anomalous as a whole, but not necessarily as individuals (Chandola et al., 2009).

Most of the anomalies represent information that is usually critical for the system. For example, a telemetry anomaly may indicate a

\* Corresponding author.

E-mail address: [gonzalo.farias@pucv.cl](mailto:gonzalo.farias@pucv.cl) (G. Farias).

sensor failure or poor system performance (Abdelghafar et al., 2019). Since telemetry data is naturally sequential, the use of time-series methods is suitable in this context. Previously developed works on anomaly detection include support vector machines and similarity query methods (Farias et al., 2006), discrete encoding and string comparisons (Dormido-Canto et al., 2006), using artificial neural networks such as Autoencoder (AE) (Farias et al., 2020a) and recurrent neural networks (RNN) (Farias et al., 2020b) to find anomalous patterns in discharges related to plasma behaviours in nuclear fusion, or the usage of convolutional neural networks (CNN) to detect exoplanets candidates from observed light-curves (Cuéllar et al., 2022b).

Various methods designed for signal processing and feature extraction in time series have been developed as anomaly detection systems. For example, the discrete wavelet transform, which provides accurate time-localized frequency information, is combined with Bayesian analysis to detect smooth and abrupt changes in variance and frequency in the approach proposed by Alarcon-Aquino and Barria (2001). Similarly, the work proposed in Thill et al. (2017) combines the discrete wavelet transform with maximum likelihood estimation for unsupervised anomaly detection. Other approaches to anomaly detection have used fluctuation analysis to identify anomalies. The work presented on Du et al. (2023) achieves a low linear time complexity with a system that computes the difference between the fluctuation of an object and its neighbours, and identifies the object with the larger difference as an anomaly. On the other hand, the approach on Gao et al. (2020) proposes a framework for anomaly detection based on time series decomposition and convolutional neural networks. Time series decomposition handles effectively complex patterns and simplifies the architecture of the network.

The first and most common method to detect abnormal patterns in telemetry data can be considered to be the out-of-limits (OOL) alarm, implemented on numerous ESA missions (Martínez-Heras and Donati, 2014). It is based on establishing thresholds for measurements using traditional statistical methods, and an alarm is activated if a measurement exceeds these thresholds. However, this method is not suitable for large-scale sequences due to its time consumption and lack of ability to deal with failures and abrupt operating conditions. In addition, there are some types of anomalies, such as contextual ones, whose characteristics do not exceed the threshold so they would not be detected.

Indeed, the analysis and classification of this large amount of data generates a significant human cost in the mission, even if the transmission is carried out in parts. The increasing development of machine learning techniques allows the implementation of new approaches for the automatic detection of anomalies in special missions that facilitate this task. The work presented in Fujimaki et al. (2005) proposes an anomaly detection system based on Kernel Feature Space. A model is built that represents the behaviour of past nominal data and forecasts the current state by reporting unexpected changes in a nonlinear feature space generated by the kernel. Other approaches use principal component analysis (PCA) to address the problem of high dimensionality and later a binary support vector machine (SVM) classifier for fault detection such as Yairi et al. (2010), Nassar et al. (2015), Gao et al. (2012) and Xiong et al. (2011).

Given the scarcity of labelled data sets in this area, working with unbalanced data is one of the challenges in anomaly detection (Chandola et al., 2009). The literature has focused on developing unsupervised approaches based on prediction errors, whose objective is to identify a model that learns the normal behaviour of a time series and predicts the following samples, detecting an anomaly when the prediction error exceeds a certain threshold. The most common prediction approaches are based on Gaussian regression and relevance vector autoregressive model (RVM) (Pang et al., 2018, 2019), artificial neural networks (ANN) (Fernández et al., 2017), Autoencoder (AE) (Sakurada and Yairi, 2014; OMeara et al., 2018), Variational Autoencoder (VAE) (Ahn et al., 2020; Memarzadeh et al., 2020; Tariq et al., 2019), recurrent neural

networks (RNN) (Hundman et al., 2018; Li et al., 2019) and some deep novel techniques such as Transformer (Meng et al., 2020) and Generative Adversarial Networks (GAN) (Geiger et al., 2020; Song et al., 2020; Yu et al., 2021).

However, these unsupervised methods applied to spacecraft telemetry present a series of disadvantages such as: first, due to the nature of the approaches themselves, they do not always include prior knowledge of well-known anomalies and are based on an assumed distribution of anomalies. Second, these approaches are not designed for a specific type of anomalies so, they may work with point anomalies (change point detection) (Tartakovsky et al., 2012; Aminikhanghahi and Cook, 2017) but not with contextual or collective anomalies. Finally, the complexity of the models makes it difficult to get an idea of the possible cause of the anomaly. The latter is key to avoid or mitigate a major failure.

The possible explanation of the causes of an anomaly is valuable information that can allow system failures to be avoided in the future. In this sense, explainable machine learning is a field of explainable artificial intelligence (XAI) that seeks to provide clear explanations of how machine learning algorithms make decisions. The goal of XAI in spacecraft anomaly detection is to provide mission operators and spacecraft engineers with a clear understanding of why a particular anomaly was detected, and what factors contributed to that detection. This makes it a very useful tool for improving spacecraft safety, helping mission operators quickly identify and respond to anomalies and reducing the risk of mission failure (Das and Rad, 2020).

To contribute to improving the reliability and safety of space missions, this work proposes a supervised anomaly detection system designed for five different types of anomalies. This approach has two stages, the first is a feature extraction process designed to differentiate abnormal from nominal behaviour in different domains (magnitude, frequency and wave-form). The second stage is the construction of the machine learning model from the features extracted to perform a binary classification task and thus predict anomalous behaviours. Different machine learning algorithms have been applied and compared. The detection system is evaluated on data not seen by the model from two databases of real spacecraft missions. The main contributions of this paper can be summarized as follows:

- The analysis of five anomaly types (Type I, Type II, Type III, Type IV, and Type V) present in two different space mission datasets of NASA, namely, SMAP and MSL. This study expands the limited set of point, and contextual and collective anomalies to also incorporate significant changes in magnitude with Type II and V anomalies, waveform with Type III anomaly, and frequency with Type IV anomaly over an entire time interval (See Section 3.2 for more details).
- A supervised system has been designed for anomaly detection by extracting features based on changes in the magnitude and frequency of known anomalies. Three feature extraction techniques have been used: LSTM, STFT and moving average.
- Classification models have been obtained with Adaptive Boosting, Random Forest, Multilayer Perceptron Neural Networks, Support Vector Machines and K-Nearest Neighbours. Results have been evaluated in terms of precision and a false positive mitigation stage has been added. They have also been compared with recent research that has worked with the same databases.
- An explainability analysis of the proposed anomaly detection system has been developed, which provides added value.

This article is developed in the following sections: The second section presents various proposals for spacecraft anomaly detection approaches found in the literature and briefly describes the benchmark of our work. The analysis of types of anomalies present in two real space mission data sets is described in the third section. The fourth section contains the proposed methodology for the anomaly detection system. The fifth section presents and discusses the results obtained

with the different machine learning models for various metrics. Finally, the sixth section presents the research conclusions, highlights, and prospects.

## 2. Related works

As mentioned above, detecting abnormal patterns in telemetry data has gained significant prominence recently. Due to the immense total of telemetry measurements that are the mainstay of the operations and health satellite monitoring (Pang et al., 2019), the quickest way to assess it is using Machine Learning (ML) based methods. Extensive research has been conducted and documented in the literature, emphasizing the application of ML techniques for anomaly detection on spacecraft telemetry. A discussion of the relevant works related to ours is presented in this section.

Table 1 summarizes the papers included in this short review. The space mission is presented in the first column as whether the dataset used is available to the public, i.e. there is a public dataset with labelled anomalies for the telemetry of this mission. The third column presents the machine-learning technique used for feature extraction and anomaly detection. The fourth column contains references to the articles in the bibliography that use this approach. Finally the fifth column shows the best performance metric achieved, the work where it is achieved is highlighted in bold.

Most of the anomaly detection anomalies are designed and tested on a simulated spacecraft telemetry database. The work presented on (Sakurada and Yairi, 2014) reduces dimensionality by using an AE on synthetic channels and comparing it with kernel PCA (kPCA) obtaining an AUC of 0.8852 for PCA, 0.9354 for AE and 0.8862 for kPCA. In Nassar et al. (2015), Fuertes et al. (2016) fault detection approaches also based on PCA are presented. The first one uses the multivariate projection technique and is implemented on spacecraft ADCS management operations to overcome faulty states. The second paper contains a decision frontier fitted by a One-Class Support Vector Machine. The NOSTRADAMUS software implements this method.

On (Liu et al., 2017) a strategy for detecting anomalies is proposed that studies the uncertainty estimation of LS-SVM along with Statistical Analysis, and it is tested with real and simulated data obtaining an accuracy of 95.8%. However, the majority of the parameters are determined through experiments and the expertise of professionals. In Ahn et al. (2020) the use of generation models such as VAE and GAN is proposed. A Bayesian optimization algorithm is used to optimize the network hyperparameters. Wang et al. (2022) proposes the method by Deviation Divide Mean over Neighbours (DDMN) that mitigates false positives due to errors on the telemetry channels. Then LSTM is used to model the data waveform.

There are approaches designed for particular missions, for example, the work presented on Codetta-Raiteri and Portinale (2014) claims that Dynamic Bayesian Networks can be a suitable modelling and inference method to detect, identify and recover from faults on autonomous spacecraft. This method has also been applied to ARPHA, an onboard software by the ESA. In Biswas et al. (2016), unsupervised clustering is implemented on the telemetry divided into temporal windows for NASA Lunar Explorer LADEE. Zheng et al. (2016) proposes using the State-counting method (SCM) to measure changes in the channels as variation features. The Sequential Probability Ratio Test (SPRT) is introduced to conduct anomaly detection and applied to telemetry from China Xi'an Satellite Control Center (XSCC) obtaining an accuracy of 99.63%. For JAXA satellite SDS-4, the authors in Yairi et al. (2017) propose health monitoring and detection of anomalies using reduction of dimensionality and clustering. Pilastre et al. (2020) proposes the Convolutional Anomaly Detection Strategy based on dictionary decomposition (C-ADDICT) that transforms each input signal by encoding it through joint activation and uniform combinations of filters. This enables the capture of correlations across the input signals. The work presented in Carlton et al. (2018) presents a telemetry fault detection

system based on transient and event detection with a mean average window. The method evaluates the telemetry series comparing it with the local absolute value. This approach can be implemented in any spacecraft architecture since there is no dependency on specific parameters from the mission, evaluated on GEO ComSat telemetry.

Other missions such as the NASA shuttle valve, Korea MultiPurpose Satellite2 (KOMPSAT-2) and the Tsinghua University smart communication satellite have been used to study telemetry anomaly detection approaches; for example, the use of Grey Wolf Optimization algorithm (GWO) with Extreme Learning Machine (ELM) to comparing the prediction error to detect anomalous events (Abdelghafar et al., 2019) and PCA with LSTM for the improvement the anomaly detection performance (Tariq et al., 2019). In addition, the work proposed by Obied et al. (2023) presents the first deep clustering task applied to spacecraft telemetry and that uses a dynamically weighted loss function. This approach helps to discover and extract realistic and reasonable hidden patterns from the normal operational data from the first Slovak satellite named skCube.

### 2.1. Base-line approaches

The only available dataset of telemetry with known anomalies labelled comes from a public repository from NASA and corresponds to the missions Mars Science Laboratory (MSL) and Soil Moisture Active Passive (SMAP). The first approach applied to this dataset which is the first benchmark for this article is presented in Hundman et al. (2018). It consists of an RNN-based and dynamic threshold approach for anomaly detection. Using the LSTM recurrent neural network for telemetry forecasting, a smoothed prediction error is calculated; then, using a dynamic threshold the sequences are classified as anomaly or nominal, and finally, a false positive mitigation strategy is developed to improve the detection rates, obtaining a  $F_{0.5}$  score of 85.5% for SMAP and 86.7% for MSL. This approach led to the development of several works, for example, Wu et al. (2018) that presents Hierarchical Temporal Memory (HTM), HTM is well-suited for handling dynamic patterns in data and possesses the ability to use past information, resulting in higher precision predictions. HTM adapts quickly to data achieving an F1 score of 73.5%. Additionally, the authors in Li et al. (2019) present a framework that stacks three LSTM-based predictors, a Support Vector Machine (SVM) based predictor and a fully connected layer, then the error distribution is estimated using KDE, and the dynamic threshold algorithm from Hundman et al. (2018) is applied to get abnormal patterns in the data. This method was tested obtaining a recall of 86.1% and 91.3% and a precision of 81.6% and 90.0% respectively. More advanced techniques are applied in other articles such as Geiger et al. (2020), where the authors propose TadGAN, which is an anomaly detection development built on Generative Adversarial Networks (GANs). The generator and discriminator are based on LSTM RNN architectures. The test databases include NASA's MSL and SMAP obtaining a total F1 score of 62.0%. Similarly, the authors in Xiang and Lin (2021) combined RNN, EVT and GRU. The prediction error is calculated using an Exponentially Weighted Moving Average (EWMA). The detection rule is automatically set through EVT. This approach was tested obtaining 87.4% and 91.2% of F1-score respectively. Our previous work (Cuéllar et al., 2022a) is our second main benchmark and presents preliminary results on a supervised approach for anomaly detection in spacecraft telemetry, a feature extraction process is performed using three detectors based on RNN, moving average (MA) and Fourier transform. A machine learning-based model was trained using the AdaBoost algorithm to perform anomaly classification. This system obtained 90.9% precision and 60.6% precision in SMAP missions. The most recent approaches include the MAG method presented on Yu et al. (2023) which employs a graph neural network with embedding vectors to characterize inherent attributes in each dimension and an attention mechanism to identify short-term interactions across dimensions joined with a LSM network for temporal features obtaining a precision of

**Table 1**

Anomaly detection in spacecraft telemetry approaches in the literature. The best performance metric per mission is presented and the work where it is achieved along with the detection method used are highlighted in bold.

Mission	Detection method	References	Best results
Simulated	MPPCA & K-Means, KPCA, LS-SVM, SVM & ANN, Autoencoder, PCA OC-SVM, SVM, GPR & RVM, DLSR, VAE& GAN, <b>DDMN-LSTM</b>	Fujimaki et al. (2005), Yairi et al. (2010), Xiong et al. (2011), Gao et al. (2012), Sakurada and Yairi (2014), Nassar et al. (2015), Fuertes et al. (2016), Liu et al. (2017), Pang et al. (2019), Pilastre et al. (2020), Ahn et al. (2020), Wang et al. (2022)	Precision: 99.9% Recall: 82.5% F1: 90.4%
Real(no named)	ANN, LS-SVM, AE & LSTM, GR & RVM, HAC, PseudoPeriod GNN-DTAN, <b>MSD-BiLSTM</b>  DCDSPOT-NETE	Fernández et al. (2017), Liu et al. (2017), OMeara et al. (2018), Pang et al. (2018), Zhang et al. (2020), Jiang et al. (2019), Xie et al. (2021), Chen et al. (2021), Zeng et al. (2022)	Accuracy: 96.3% Precision: 97.9% Recall: 98.3% F1: 98.1%
ExoMars rover	DBN	Codetta-Raiteri and Portinale (2014)	
XMM	OOL alarm	Martínez-Heras and Donati (2014)	
LADEE	UPGMA CLustering	Biswas et al. (2016)	
XSCC	<b>SPRT</b>	Zheng et al. (2016)	Accuracy: 99.6%
JAXA SDS4	MPPCACD & OCSVM & CDL	Yairi et al. (2017), Pilastre et al. (2020)	
GEOComSat	Tukey Method	Carlton et al. (2018)	
NASA Shuttle Valve	<b>GWO-ELM</b>	Abdelghafar et al. (2019)	Precision: 90.6% Recall: 98.3% F1: 94.3%
KOMPSAT-2	<b>LSTM-PCA</b>	Tariq et al. (2019)	Precision: 66.6% Recall: 90.9% F1: 76.9%
Tsighua University	CP-WSO	Wan et al. (2019)	
NASA FOQA	<b>CVAE</b>	Memarzadeh et al. (2020)	Precision: 36.8% Recall: 27.3%
SK Cube	<b>Deep Clustering LOP</b>	Obied et al. (2023)	Precision: 97.3% Recall: 98.1%
MSL and SMAP (Available)	LSTM, HTM, LSTM & SVM & FCNN, OC-SVM & LSTM Bi-LSTM, Transformer, GAN, GRU & EVT, GAN TL, AdaBoost <b>MAG</b> , TCN	Hundman et al. (2018), Wu et al. (2018), Li et al. (2019), Wu et al. (2020), Pan et al. (2020), Meng et al. (2020), Geiger et al. (2020), Song et al. (2020), Xiang and Lin (2021), Yu et al. (2021), Baireddy et al. (2021), Cuéllar et al. (2022a), Yu et al. (2023), Liu et al. (2023)	Precision: 95.4% Recall: 99.2% F1: 97.2%

95.26% in SMAP and 95.94% in MSL. Finally, the method proposed on Liu et al. (2023) that employs dynamic graph attention to model the data behaviour combined with temporal convolution networks to extract multidimensional features, then a threshold is used to detect anomalies present on the data, obtaining a recall of 98.15% on MSL and 90.19% in SMAP.

### 3. Known anomalies from SMAP & MSL databases

#### 3.1. Data description

The dataset used comes from a public repository from Hundman et al. (2018) which contains spacecraft telemetry data from the NASA missions SMAP and MSL. The available data consists of measurements of 55 and 27 telemetry channels, each one associated with one of the 12 and 27 modules of SMAP and MSL, respectively. For example, from channels T-1 (SMAP) and T-8 (MSL), the time series of Fig. 1 correspond to measurements of variables on the temperature module.

The database maintains the sequential nature of each channel despite the sample time being unknown, so all channels have different

lengths. Besides that, there is no record of simultaneous measurements, so there is no explicit correlation between telemetry channels.

Each channel  $x$  has  $n \times m$  size. The samples  $x[t]$  with  $t = 0.1, \dots, n$  are  $m$ -size vectors whose first element corresponds to the measurement or the telemetry value. The rest of the elements are associated with commands sent and received by the spacecraft modules. Thus, for each module, there are 2 elements on  $x[t]$  whose values are 1 if there is a command sent and received and 0 otherwise. In this way  $m = (\text{telemetryvalue}) + (\text{numberofmodules}) * (\text{sentandreceived})$  i.e.  $m = 1 + (12 * 2)$  for SMAP and  $m = 1 + (27 * 2)$  for MSL.

Despite the nature of the work presented in Hundman et al. (2018) is unsupervised, each channel contains a training series that corresponds to the measurement of nominal behaviour and a test series that contains different anomalous sequences labelled by experts from NASA's Jet Propulsion Laboratory (See Fig. 1). For example on channel T-1 of SMAP there are labelled anomalies on intervals [2399, 3898] and [6550, 6585]. Even when there are only two labelled anomalies on this channel there are 1499 and 35 labelled anomalous samples respectively. In total, there are 66 anomalies in SMAP, and 35 on MSL that correspond to 75387 anomalous samples, i.e. 16.12% of the total



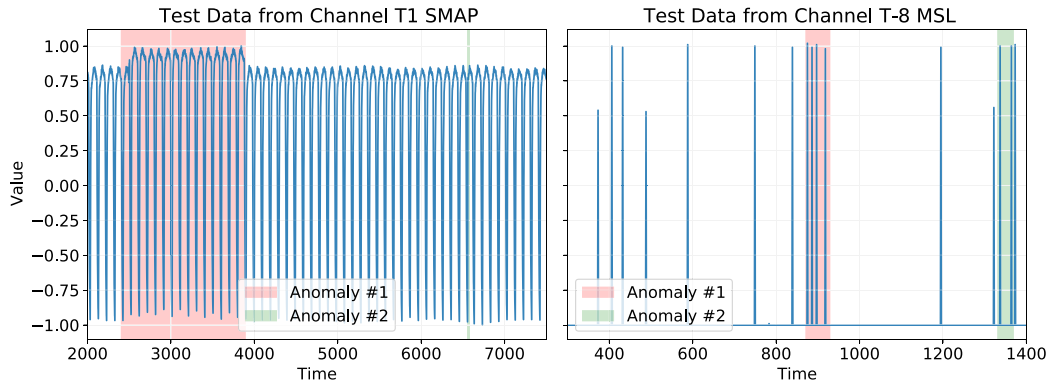


Fig. 1. Example telemetry data from NASA's public dataset. On the left, channel T-1 from SMAP, at samples [2399,3898] and [6550,6585] contains labelled anomalies. On the right, channel T-8 from MSL, at samples [870,930] and [1330,1370] contains labelled anomalies.

Table 2

NASA spacecraft telemetry public database description.

	SMAP	MSL	Total
Number of channels	55	27	82
Number of labelled anomalies	66	35	101
Number of mission modules	12	27	–
% anomalous samples	–	–	16.12%
Number of telemetry samples evaluated	–	–	467724

samples from the telemetry channels 467724. There may have been labelling errors made by human judgement; however, this classification will be taken by ground truth in this work. Table 2 shows the details from the database used.

### 3.2. Types of anomalies

Telemetry data represent spacecraft measurements expressed as time series, which can have a continuous or discrete nature. The labelled NASA dataset does not have a distinction between time-series types, and the approaches described in the literature applied to it only consider point and contextual anomaly types. To understand the conditions and forms of the anomalies that occur in spacecraft telemetry, we performed a study of the known anomalies from the SMAP and MSL missions. The original set of 82 telemetry channels was divided into 101-time series in such a way that each one has one labelled anomaly.

For this document, we defined and established the following types of time series: periodic, binary, and non-stationary value time series. Periodic, also known as stationary, are those continuous or discrete nature time series that have a repeating pattern on a time interval. Binary refers to a discrete nature time series that takes only two values,  $-1$  and  $1$ . Finally, non-stationary refer to time series that do not have a repeating pattern.

The 101 series dataset of two different missions contains 15 binary, 28 periodic and 58 continuous value time series. The context of each type of signal is different and anomalies can occur under different conditions in the time domain; therefore, we have identified five types of anomalies and methodologies for detecting them (See Fig. 2):

- Type I: A value that lies outside the expected range, occurs in all types of time series and is detectable by evaluating the distance from its neighbour's values. These belong to the set of point anomalies.
- Type II: The main variation lies in the magnitude over an interval, there are no relevant wave-form nor frequency changes. This anomaly occurs on periodic and continuous values time series and is detectable by weighting its neighbourhood. These belong to the set of collective anomalies.

Table 3

Identified anomalies on a labelled dataset from SMAP and MSL missions.

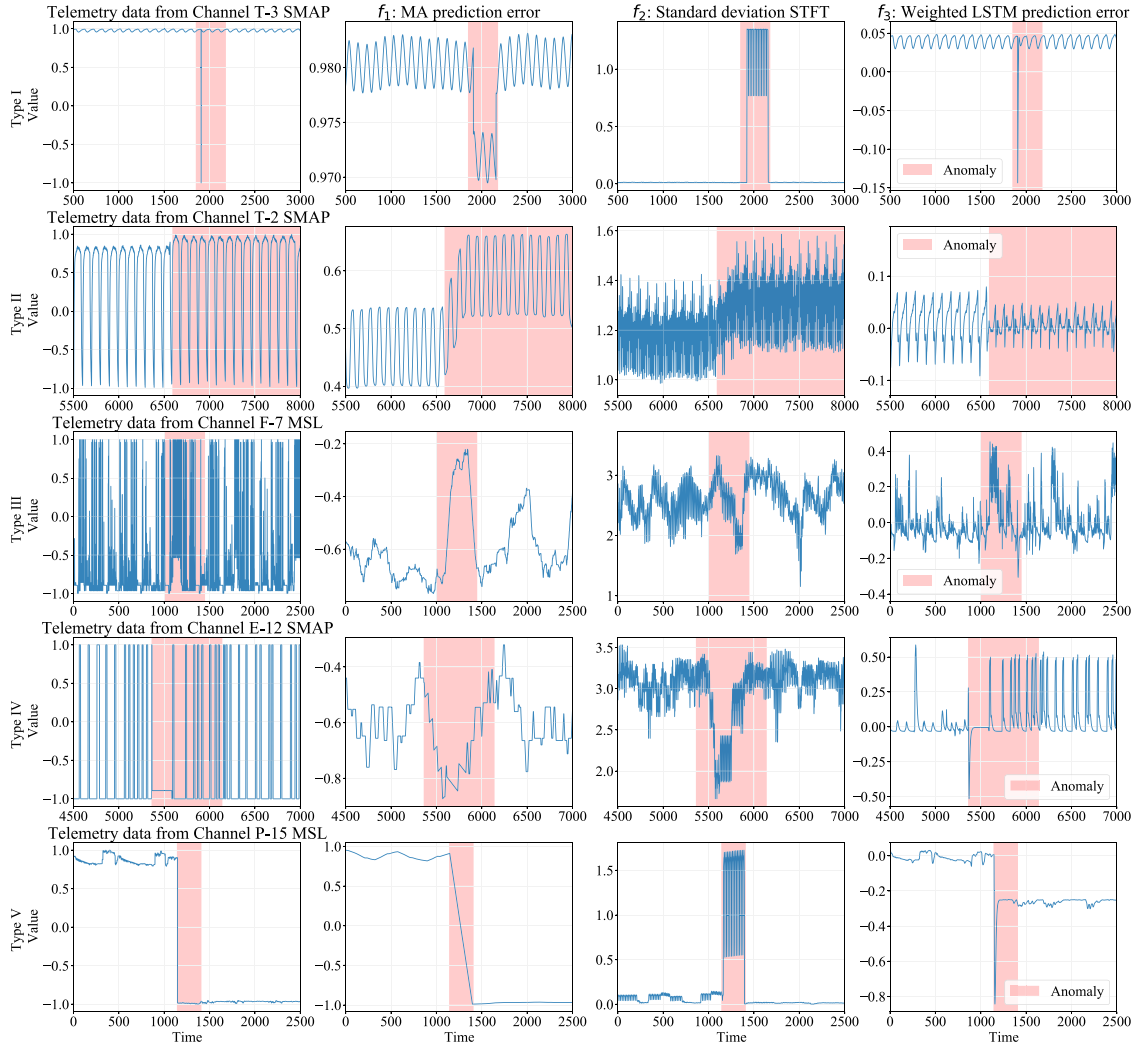
Mission	Type I	Type II	Type III	Type IV	Type V
SMAP	13	23	17	6	13
MSL	3	22	8	–	3
Total	16	55	25	6	16

- Type III: There is a significant variation in the amplitude, but the values are not abnormal, so there is an alteration in the waveform of the time series. This anomaly may occur on periodic and continuous values time series and is detectable by calculating the difference from the expected values. These belong to the collective anomalies set.
- Type IV: The time-series values are maintained over the interval; however, there is a change in the frequency. This anomaly occurs on binary and periodic time series and can be detected by comparing the spectral components over the interval. These belong to the contextual anomalies set.
- Type V: The abnormal pattern corresponds to the abrupt variation of magnitude values, but these values are not anomalous by themselves. This anomaly can occur on continuous values series and is detected by calculating the prediction error with the expected values. This type belongs to the contextual anomalies set.

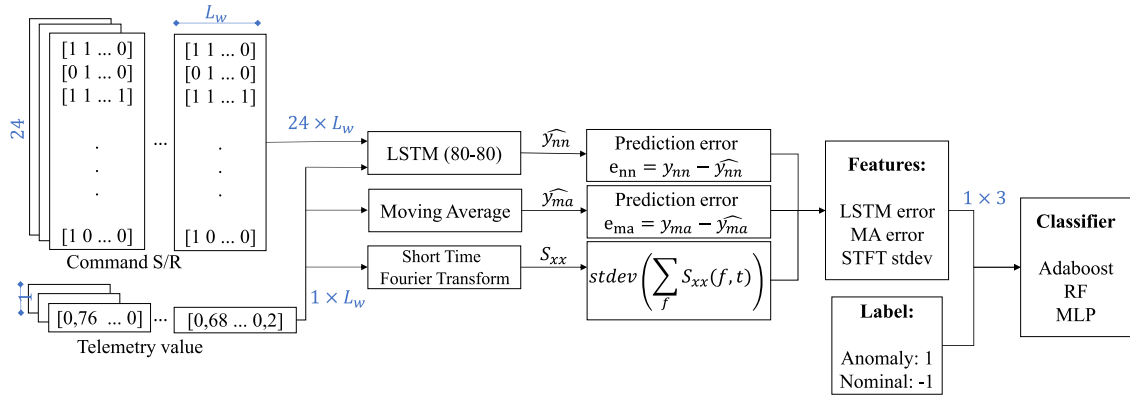
Table 3 shows the number of anomalies by type for each spacecraft mission. It must be highlighted that anomalies that can be detected with prediction and distance-based methods used in the majority of approaches in the literature, i.e., Types I and III, represent only 40.5% of the total. Since each identified type of anomaly has its own detection methodology, it is required to design a system that contemplates all the scenarios in which they may occur.

## 4. Proposed approach

The proposed methodology presented in the workflow from Fig. 3 consists of two main stages. The first is the feature extraction process to generate a feature vector that contains knowledge about magnitude, frequency and waveform changes on telemetry channels. This feature extraction process was designed to differentiate the abnormal from the nominal behaviour in different domains. The second stage is the obtaining of the machine learning model from the features extracted, the description of different algorithms in the task of binary classification is presented in this section. This section also presents the anomaly selection process performed with the classifier output and the performance metrics used to evaluate the proposed approach.



**Fig. 2.** Types of anomalies identified and feature extraction results. Each row contains an example of one of the five types of anomalies identified (Type I, Type II, Type III, Type IV, and Type V). The first column corresponds to the original time-series values for the telemetry channel. The second column presents the moving average prediction error ( $f_1$ ). The standard deviation of STFT ( $f_2$ ) is on the third column. Finally, the fourth column shows the smoothed LSTM prediction error ( $f_3$ ).



**Fig. 3.** Workflow of the proposed approach.

#### 4.1. Feature extraction

In order to detect the types of anomaly identified from known anomalies, we generate a feature vector that contains the following three features:

- $f_1$ : A feature that brings knowledge about waveform changes (Types I, III and V); an LSTM will be used for this goal.
- $f_2$ : A feature that includes a time-series spectral analysis to detect abnormal changes in frequency (Type IV), Short Time Fourier Transform will be applied.
- $f_3$ : A feature that captures weighting information in a window to detect abnormal changes in magnitude (Type II).

Fig. 2 shows examples of the feature extraction results for examples of the five types of anomalies identified. Each feature highlights a specific behaviour within the time series, and their combination facilitates the easy detection of all identified types of anomalies.

##### 4.1.1. Moving average predictor (MA)

A moving average predictor is a statistical technique used to forecast future values based on past observations of a time series. The predicted sample is equal to the weighted average between the past  $n$  samples. One of the advantages of using a moving average predictor is that it enables the identification of time trends by smoothing out alterations on series. However, it can also lead to a delay in detecting sudden changes or shifts in the data (Hansun, 2013). For the purpose of this work, each sample will have the same weight. Eq. (1) shows the calculation of the predicted sample  $\hat{y}_{ma}(t)$  where  $x$  denotes the series value and  $n$  is the number of weighted samples, i.e., the window size.

$$\hat{y}_{ma}(t) = \frac{x(t) + x(t-1) + \dots + x(t-(n-1))}{n} \quad (1)$$

When a value  $\hat{y}_{ma}(t)$  is predicted for a sample  $t$ , we calculate the prediction error using  $e_{ma}(t) = |y_{ma}(t) - \hat{y}_{ma}(t)|$ , where  $y_{ma}(t) = x(t+1)$ . The prediction error  $e_{ma}(t)$  will be the first feature  $f_1$  of the proposed system.

##### 4.1.2. Short-time fourier transform (STFT)

Is a variation of the Fourier transform used in signal processing to examine the frequency of a signal as it evolves. The STFT applies the Fourier transform separately to time segments or windows from the signal; this allows the inspection of the behaviour of a localized frequency over time. Eq. (2) shows the calculation of the STFT on a signal  $s(t)$  using a window function  $w(t)$ . The magnitude of the STFT  $S_{xx} = |STFT|$  is known as the spectrogram, which is a 2D matrix that shows the amount of the frequency component of a signal (y-axis) respect time (x-axis).

$$STFT(t, \omega) = \int s(t')w(t' - t)e^{-j\omega t'} dt' \quad (2)$$

However, the STFT has limitations on time and frequency resolution. The size of the window used for analysis determines the trade between time and frequency resolution. A smaller window size gives better time resolution but poorer frequency resolution and vice-versa. Therefore, the choice of window size is critical for accurate spectral analysis using STFT. Additionally, the STFT assumes that the signal is stationary within each window, which may not be true for some signals (Chen and Ling, 2002).

Since there is a time-frequency matrix for each window analysed and we need to extract one feature, all frequency components for an instant of time will be added. Then the standard deviation of the window will be calculated to see how the frequency changes per window. Therefore, the second feature will be calculated with Eq. (3)

$$f_2 = \text{stdev} \left( \sum_f S_{xx}(f, t) \right) \quad (3)$$

##### 4.1.3. LSTM predictor

LSTM is a powerful deep-learning model that can capture complex patterns in time series data and make accurate forecasts. It is designed to overcome the limitations of traditional RNNs in capturing long-term dependencies in sequential data.

LSTM has a special memory cell that allows it to selectively retain or forget information over time. The input, output and forget gates control the memory cell. These gates are controlled by sigmoid activation functions that determine how much information should be passed through the cell. During training, LSTM learns to adjust the weights of these gates according to the time-series patterns. This allows the model to remember important information over long periods and discard irrelevant information. LSTM is particularly useful for time series forecasting because it can handle sequences of varying lengths and can predict multiple time steps ahead. It is also robust to noisy data and can handle missing values (Malhotra et al., 2015).

The prediction is performed as follows. A telemetry channel contains a multidimensional time series  $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots, \mathbf{x}^{(n)}\}$ , each sample  $\mathbf{x}^{(i)}$  in the time-series is a  $m$ -dimensional vector  $\{x_1^i, x_2^i, \dots, x_m^i\}$  which elements are the input variables. The value of  $l_s$  determines the input size of the model for each  $\mathbf{x}^{(i)}$ , the prediction length  $l_p$  determines how many samples ahead will be estimated, and  $d$  ( $1 \leq d \leq m$ ) is the size of the prediction. Since we want to predict only the next sample for a given channel  $d = 1$  (Li et al., 2019). In this work, the inputs are vectors of  $L_w$  telemetry values and the respective codified commands, that means, for SMAP  $\mathbf{x}^{(i)}$  is a vector with size  $1 \times 25 \times L_w$  and for MSL  $\mathbf{x}^{(i)}$  is a vector with size  $1 \times 55 \times L_w$ .

When a value  $\hat{y}_{nn}^t$  is predicted for each window  $t$ , the prediction error  $e_{nn}^{(t)} = |y_{nn}^t - \hat{y}_{nn}^t|$  is calculated, with  $y_{nn}^t = x^{(t+1)}$ . Following the approach presented in Hundman et al. (2018), the errors are then smoothed to reduce abrupt changes that frequently occur with LSTM prediction using exponentially weighted average (EWMA). The smoothed prediction error  $e_{smn}^{(t)}$  will be the third feature  $f_3$  of the proposed system.

#### 4.2. Classification models

The proposed method will be evaluated using the same methodology described in Hundman et al. (2018). For the experiments, a training set consisting of 70% of the labelled anomalies was used, while the remaining 30% was allocated for testing purposes. To evaluate the generalization, cross-validation was performed with the following machine learning-based classifiers.

##### 4.2.1. Adaptive boosting (AdaBoost)

Boosting is a method that aims to Enhance the efficacy of any supervised learning algorithm by combining the results of several weak or baseline classifiers to obtain a robust final classifier. One of the most popular Boosting techniques is the Adaptive Boosting algorithm (AdaBoost) (Schapire, 2013), which will be used in this proposal. This algorithm, by iterative training of the weak or base classifiers, assigns greater importance to the previously misclassified data and obtains a new classifier. In this way, it manages to adapt and obtain better results by increasing the accuracy of the algorithm.

##### 4.2.2. Random forest (RF)

Random Forest is one of the top-performing classifier methods, consisting of an ensemble of decision trees that has randomized inputs taken by bootstrapping samples from the train set, as in standard bagging (James et al., 2017). The trees are built by enforcing a rule where, at each node, only a randomly selected subset of attributes is eligible for splitting. The subset size is typically relatively small, often equivalent to the square root of the total attribute amount. Additionally, the trees are constructed without employing any pruning techniques.



index	0	1	2	3	4	5	6	7	8	9	10
label	-1	-1	1	1	1	-1	-1	1	-1	-1	-1

Anomaly 1
Anomaly 2

index	0	1	2	3	4	5	6	7	8	9	10
pred	-1	1	1	-1	-1	-1	-1	-1	-1	1	1

Fig. 4. Example of selection anomaly criteria. The true anomalous sequences are marked in green and the predicted anomalous sequences are marked in yellow. index, label and pred correspond to the sample index, the true label and the predicted label, respectively. In this scenario, the first anomaly predicted is recorded as True Positive, since there is an overlap on sample 2; a False Negative is recorded since Anomaly 2 is not detected by the predictor; finally, the second prediction is recorded as False Positive since there is no overlap between it and a labelled anomalous sequence.

#### 4.2.3. Multilayer perceptron

Is a type of artificial neural network conformed by three types of layers: input, output and hidden. The data to process is entered by the input layer. A certain amount of hidden layers are added between the input and output layers to extract further information from the data. The output layer performs the classification task. The backpropagation algorithm is used to train all the neurons within. This method is designed specifically to address non-linearly separable problems by approximating any continuous function (Murtagh, 1991).

#### 4.2.4. Support vector machines

A support vector machine (SVM) classifier is a machine learning algorithm used for binary classification problems. It is designed to find a hyperplane i.e. a linear decision boundary in a high-dimensional feature space, that separates the input vectors into two classes. The SVM classifier has a great ability to generalize well to unseen data. This is achieved by maximizing the margin, which is the distance between the decision boundary and the closest data points from each class (Cortes and Vapnik, 1995).

#### 4.2.5. K-nearest neighbours

Is a supervised machine learning algorithm used for classification and regression problems. Given a positive integer  $K$  and an instance of data, the KNN algorithm first identifies the  $K$  points in the training data that are close to the instance (the  $K$  closest neighbours), and then it estimates the conditional probability for the neighbour's classes. Finally, KNN applies the Bayes rule to classify the instance to the higher probability class. (James et al., 2017)

### 4.3. Anomaly selection

The output classifier is a time series where each element is 1 if an anomaly is detected on the respective window and -1 otherwise. In order to use the same metrics as Hundman et al. (2018), the prediction is transformed into a set of sequences with the beginning and end of the anomaly. A true positive (TP) is recorded if the labelled and predicted sequences are overlapped at least in one sample. A false positive (FP) is recorded when a labelled sequence is not overlapped with any of the predicted sequences. Finally, a false negative (FN) is recorded when there is no overlap between a predicted sequence and an anomaly-labelled sequence. A visualization example of this selection criteria is presented in Fig. 4.

#### 4.4. Performance metrics

To evaluate the performance of the anomaly detection model in this study, the following metrics are employed:

- Precision: The ratio of sequences classified as anomalies that are true anomalies, also known as reliability.

$$precision = \frac{TP}{TP + FP} \quad (4)$$

- True Positive Rate (TPR): The ratio of true anomalies that are classified as anomalies, also known as recall.

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

- The harmonic average of precision and recall leads to the F-score:

$$F_\beta = (1 + \beta) \times \frac{precision \times recall}{\beta^2 precision + recall} \quad (6)$$

For  $\beta = 1$ , the resulting  $F_1$  score is a classification metric that combines both precision and recall into a single measurement that gives the same weight to each metric. For  $\beta = 0.5$ , the resulting  $F_{0.5}$  score is a type of evaluation metric that strikes a balance between precision and recall by emphasizing precision over recall, meaning that it gives more weight to correct positive predictions (anomalies) while minimizing false positives (Goutte and Gaussier, 2005). In anomaly detection, it is often more important to minimize false positives (normal data points that are classified as anomalies) than false negatives (anomalies that are not detected), because false positives can lead to unnecessary and costly investigations or system shutdowns. Therefore  $F_{0.5}$  score is a better metric for anomaly detection because it helps to minimize false positives, while still detecting a high proportion of true anomalies.

## 5. Results

### 5.1. Experimental design

Following the workflow presented in Fig. 3 and to verify the performance of the proposed system, we first train an LSTM network with the same parameters used in Hundman et al. (2018) for each telemetry channel, using the 82 available training signals to learn the nominal waveform of each channel. The telemetry channels in the original test set were separated into 66 and 35-time series for SMAP and MSL respectively such that each series contains one labelled anomaly. For example, channel T-1 of SMAP has 8612 samples and contains two labelled anomalies on intervals [2399,3898] and [6550,6585]; we divided this channel into two-time series: T-1-1 with 5224 samples and a labelled anomaly on samples [2399,3898] and T-1-2 with 3388 samples and a labelled anomaly on the interval [1326,1361].

**Table 4**  
Parameter settings for the classification models.

ML model	Grid search range	Obtained parameters	
		SMAP	MSL
AdaBoost	n. estimators=(10, 500) learning rate=(0.01, 2.5)	n. estimators=50 learning rate=1 algorithm=SAMME.R	n. estimators=500 learning rate=1.5 algorithm=SAMME.R
RF	n. estimators=(100, 500) max. depth=(3, 20)	n. estimators=200 criterion=gini max. depth=16 min. samples split=2 min. samples leaf=1	n. estimators=100 criterion=gini max. depth=6 min. samples split=2 min. samples leaf=1
MLP	hidden layer sizes= ((4, 20), (4, 20), (4, 20)) activation=tanh, ReLU tolerance=0.0001, 0.05	hidden layer sizes=(15, 10, 15) max. iter=60 activation=tanh solver=adam learning rate=0.001 tolerance=0.0001 batch size=min(200,samples)	hidden layer sizes=(8, 16, 4) max. iter=60 activation=ReLU solver=Adam learning rate=0.001 tolerance=0.0001 batch size=min(200,samples)
SVM	C=(0.1, 1000) gamma=(0.1, 1000)	C=1.0 kernel=rbf gamma=0.33 max. iter.=1000	C=1.0 kernel=rbf gamma=0.33 max. iter.=1000
KNN	n.neighbours=(5,20)	n. neighbours=5 weights=uniform metric=minkowski	n. neighbours=5 weights=uniform metric=minkowski

The feature extraction step described in the previous section was performed on the 101 resulting time series for sliding window size  $L_w = 250$  and DFT points  $n_F = 20$ . Additionally, each window is labelled “1” if contains a sample of the originally labelled anomaly and “0” otherwise; on the T-1-1 example, the window from the interval [2100,2350] is labelled “1”, and the window from the interval [3900,4150] is labelled “0”. Finally, the feature values are normalized to obtain a range between 0 and 1. Using this dataset of features and labels the machine learning-based models were built and the performance was evaluated using ten rounds of cross-validation (10-fold). The adjustment parameters of the models were determined through a grid search applied to the majority, with details provided in Table 4 for the selected parameters and their corresponding grid search ranges; the remaining ones are set to default values.

## 5.2. Classification by type

The proposed method successfully detects anomalies in telemetry data. The best model obtained with the Random Forest technique ( $L_w = 250$  and  $n_F = 20$ ) records 2 false positives, 3 false negatives and 17 true positives, obtaining a recall of 85.0%, a precision of 89.5%, a  $F_1$  score of 87.2% and a  $F_{0.5}$  score of 88.5% in SMAP. In MSL the model records 2 false positives, 3 false negatives and 17 true positives, obtaining a recall of 100.0%, a precision of 84.6%, a  $F_1$  score of 89.7% and a  $F_{0.5}$  score of 87.3%.

Fig. 5 shows examples of our best classification models (See Table 6) for each one of the anomaly types identified in the previous study. Fig. 5(a) shows a time series that remains oscillating on the same value at sample 1847, where a Type I anomaly is correctly detected. This is the most common and simple anomaly to be detected because of the clear difference between the anomaly and nominal values. This is the same case as the Type V anomaly presented in 5(e). Fig. 5(b) shows an example of the detection of a Type II anomaly, this is an important challenge of the proposed system evaluation because the detectors based on predictions such as the benchmark of this work can identify only the beginning or the end of this kind of anomalies; however, the proposed method can detect the beginning of the anomaly correctly and all the way to the end of it. An example of the detection of a Type III anomaly is shown in Fig. 5(c) where a clear wave-form change of the time series appears at sample 350. The anomaly is labelled until the end of the time series but the proposed method only detects the

most visible change. This behaviour can happen due to a mislabel on the original dataset because the waveform labelled as an anomaly from samples 600 to 700 is similar to the one between samples 200 to 300 labelled as nominal. Finally, a Type IV anomaly is presented in 5(d) where a visible frequency change occurs around sample 3500, which is successfully detected, and a false positive is recorded around sample 1000.

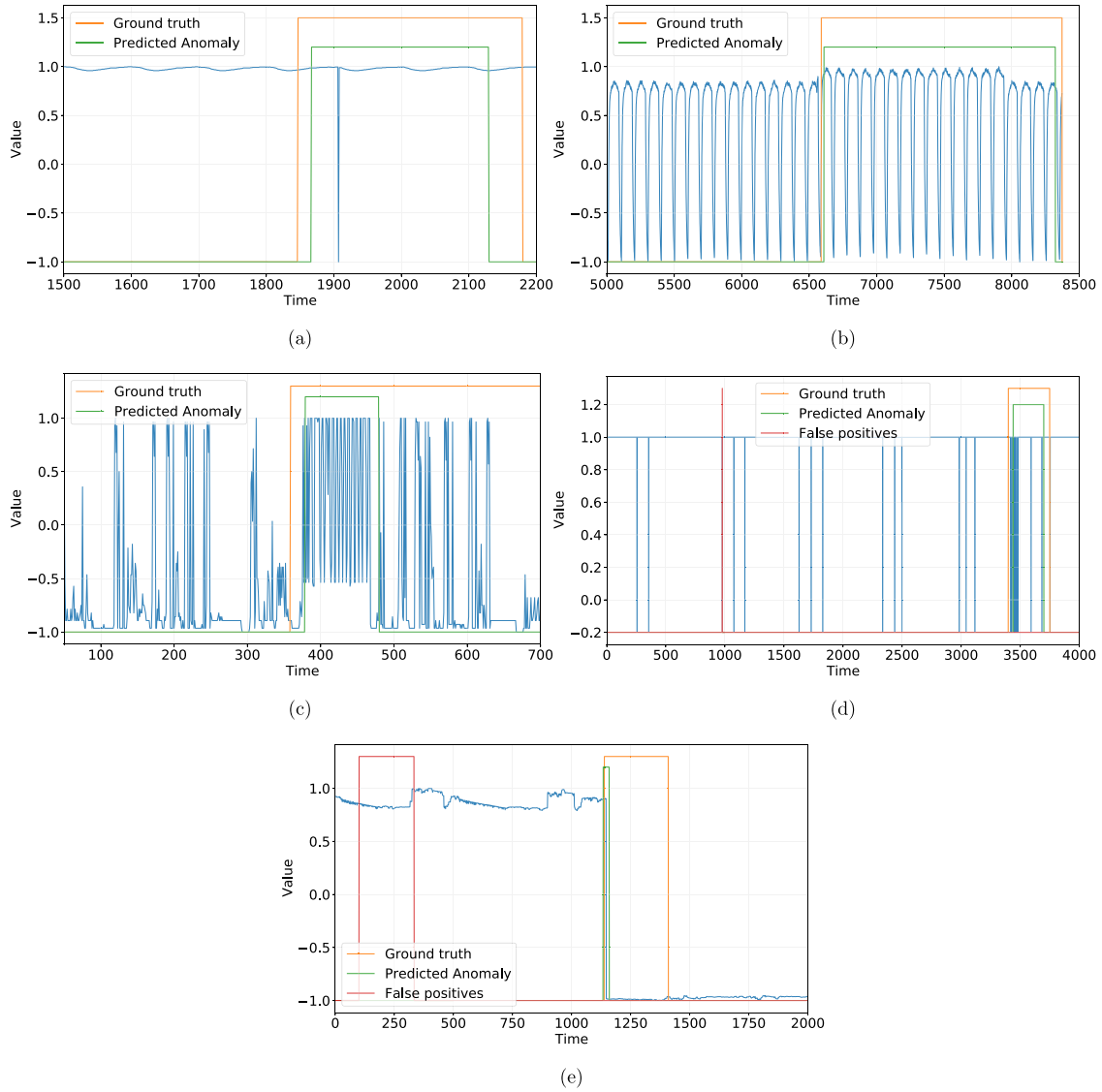
## 5.3. Performance of the proposed approach

Table 5 shows the results in mean and standard deviation for the models trained for the two missions with the 70% of the data and tested with the 30% left. The best for SMAP was the MLP with  $F_{0.5}$  of 0.540 and for MSL the AdaBoost with  $F_{0.5}$  of 0.594. The low metric values on average are due to the low number of experiments performed and an existing trade-off between Recall and Precision, where if the model classifies everything as anomaly obtains a recall of 1 but with a high number of false positives which decreases the precision. For example, the worst scenario for the Adaboost classifier was a Recall of 1.000, Precision of 0.222,  $F_1$  of 0.364 and  $F_{0.5}$  of 0.263 in MSL, where all of the anomalies were detected but 14 false positives were recorded.

The best models obtained are presented in Table 6 where the  $F_{0.5}$  score of all of them is higher than 70%. The Random forest outperforms the AdaBoost and MLP methods in terms of both Recall and precision for both missions. This is because Random forests are better at capturing non-linear relationships between the input features and the target class, especially when the non-linear relationships are complex and difficult to model. Random Forest uses decision trees with non-linear splits, whereas MLPs can only capture non-linearity through the activation functions of the neurons in the hidden layers and AdaBoost relies on linear combinations of weak classifiers. Random forests are also generally more robust to noise in the data than MLPs and AdaBoost classifiers. This is because random forests use multiple decision trees, each trained on a different subset of the features and samples, which help to reduce the impact of noisy or irrelevant features and samples on the final prediction.

## 5.4. Comparison with related work

Table 7 presents the performance metrics obtained by the proposed approach for the two missions SMAP and MSL and the comparison with



**Fig. 5.** Classification example for each anomaly type identified. The telemetry values are plotted on each figure, with the labelled anomaly in orange, the predicted anomaly in green, and the false positives in red. (a) Type I on channel T-3 of SMAP. (b) Type II on channel T-2 of SMAP. (c) Type III on channel F-7 of MSL. (d) Type IV on channel G-7 of SMAP. (e) Type V on channel P-15 of MSL.

**Table 5**

Performance (mean  $\pm$  standard deviation) obtained for each model of the anomaly detection system. The best results in terms of  $F_{0.5}$  are in bold.

		Recall	Precision	$F_1$	$F_{0.5}$
SMAP	AdaBoost	$0.886 \pm 0.162$	$0.419 \pm 0.321$	$0.504 \pm 0.275$	$0.443 \pm 0.305$
	RF	$0.850 \pm 0.096$	$0.431 \pm 0.319$	$0.522 \pm 0.256$	$0.458 \pm 0.298$
	MLP	$0.898 \pm 0.119$	$0.513 \pm 0.320$	$0.5199 \pm 0.266$	$0.540 \pm 0.302$
	SVM	$0.493 \pm 0.100$	$0.895 \pm 0.112$	$0.623 \pm 0.062$	<b><math>0.755 \pm 0.037</math></b>
	KNN	$0.886 \pm 0.113$	$0.193 \pm 0.089$	$0.306 \pm 0.120$	$0.226 \pm 0.100$
MSL	AdaBoost	$0.967 \pm 0.105$	$0.563 \pm 0.277$	$0.666 \pm 0.195$	<b><math>0.594 \pm 0.245</math></b>
	RF	$0.650 \pm 0.245$	$0.370 \pm 0.104$	$0.457 \pm 0.131$	$0.399 \pm 0.112$
	MLP	$0.292 \pm 0.081$	$0.640 \pm 0.250$	$0.384 \pm 0.072$	$0.497 \pm 0.132$
	SVM	$0.275 \pm 0.079$	$0.540 \pm 0.165$	$0.351 \pm 0.039$	$0.438 \pm 0.066$
	KNN	$0.900 \pm 0.118$	$0.195 \pm 0.086$	$0.311 \pm 0.115$	$0.229 \pm 0.096$

the baseline anomaly detection approaches from the state of the art (See Section 2.1). The metrics on percentage were obtained from the comparison performed by Bairedy et al. (2021), the validation strategy contains the anomaly selection process and the metric calculation previously described in Section 4.3 and Section 4.4 respectively. The total values of the metrics were calculated by weighting the SMAP and MSL metrics by their amount of anomalies as Eq. (7) presents,

the reason behind it is that the contribution of the performance of the models trained in a mission should be proportional to the number of anomalies present on the dataset.

$$\text{Recall}_T = 0.653\text{Recall}_{SMAP} + 0.346\text{Recall}_{MSL} \quad (7)$$

The main difference between the proposed work and the ones presented in the literature lies in two factors: First, the feature extraction

**Table 6**Best performance obtained with each ML model. Best results in terms of  $F_{0.5}$  are highlighted in bold.

	SMAP				MSL			
	Rec.	Prec.	$F_1$	$F_{0.5}$	Rec.	Prec.	$F_1$	$F_{0.5}$
AdaBoost	0.850	0.810	0.829	0.817	1.000	0.786	0.880	0.821
RF	0.850	0.895	0.872	<b>0.885</b>	1.000	0.846	0.917	<b>0.873</b>
MLP	0.800	0.762	0.780	0.769	0.667	0.727	0.696	0.714
SVM	0.350	0.583	0.438	0.515	1.000	0.423	0.595	0.478
KNN	1.000	0.303	0.465	0.352	1.000	0.393	0.564	0.447

**Table 7**

Performance results on percentage of the proposed method for each mission and comparison with some anomaly detection approaches from the literature: The LSTM-NAT based model from [Hundman et al. \(2018\)](#), the Hierarchical Temporal Memory from [Wu et al. \(2018\)](#), the combination of LSTMs and an SVM called StackedPredictor (SP-DTA) from [Li et al. \(2019\)](#), the LSTM-GAN based approach called TadGAN from [Geiger et al. \(2020\)](#), the GRU and Extreme Value Theory based approach from [Xiang and Lin \(2021\)](#) and our previous results on SMAP mission ([Cuéllar et al., 2022a](#)).

	SMAP			MSL			TOTAL		
	Prec	Rec	$F_{0.5}$	Prec	Rec	$F_{0.5}$	Prec	Rec	$F_{0.5}$
LSTM-NAT	85.5	85.5	85.5	92.6	69.4	86.8	87.9	80.0	85.9
HTM	92.7	73.9	88.2	88.2	41.7	72.1	91.2	63.0	82.7
SP-DTA	90.0	91.3	90.3	81.6	86.1	82.5	87.1	89.5	87.6
TadGAN	52.3	83.5	56.6	49.0	69.4	52.1	51.2	78.7	55.0
GRU-EVT	88.2	94.4	89.3	90.2	84.8	89.1	88.9	91.1	89.3
Previous	90.9	60.6	82.6	—	—	—	—	—	—
Our approach	95.7	100.0	96.5	94.4	100.0	95.5	<b>95.3</b>	<b>100.0</b>	<b>96.2</b>

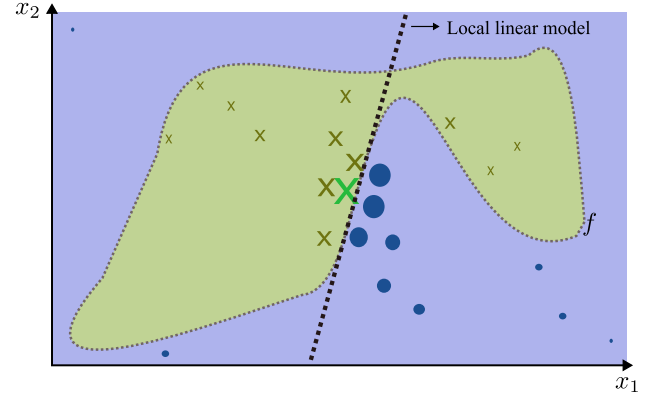
process is designed for five different types of known anomalies, and the works from the literature use unsupervised deep learning approaches to extract features so there is no prior knowledge of well-known anomalies. Second, the anomaly classification task is performed by a machine learning-based model, and its performance is evaluated on data not seen by the model so the metrics presented give a good measurement of its capacity for generalization; on the other hand, the approaches from [Table 7](#) uses the difference between the feature values and a threshold generated by an algorithm fine-tuned with the same test data. To make a fair comparison with these approaches, our parameters model was also fine-tuned with the test set. However, is important to highlight that a correct validation must be performed as shown in [Section 5.3](#).

The main difference between the proposed approach and our previous work is that the last was designed and tested for only one mission (SMAP). We also performed a fine-tuning of the parameters such as window size  $L_w$  and DFT points  $nF$  to increase the identification of anomalies in the feature extraction stage.

For SMAP our method achieves 95.7% precision and 100% of Recall, giving a  $F_{0.5}$  score of 96.5%; On the MSL dataset the precision is slightly lower, a 94.4%; it could be due to the absence of Type 4 anomalies on MSL dataset, therefore a nominal abrupt change in frequency could be wrongly classified as an anomaly. Also, our approach can be extended for its use in other missions and even applications that contain time series of the types mentioned in the previous analysis. In comparison with the related works, the proposed method obtains a higher recall which means it can correctly detect more anomalies and achieves a higher precision since it is designed for multiple types of anomalies. Therefore it can capture relevant changes in magnitude, frequency and waveform. This demonstrates the main hypothesis of this work, the inclusion of known anomalies can improve the performance of the models.

### 5.5. Explainability analysis

The results presented in the previous section show the remarkable performance of the proposed model and the literature approaches in terms of the evaluation metrics on the same dataset. However, in the anomaly detection research area, it can be challenging to understand why a particular data instance has been identified as anomalous by a machine learning model. This lack of interpretability can be a significant barrier to the adoption of machine learning-based anomaly



**Fig. 6.** Example of functioning of LIME algorithm. The decision function from the complex model  $f$  is represented by the dotted line. This complex model is a binary classifier with classes “X” (green) and “O” (purple). The instance to be explained from class “X” is highlighted in colour and size. The resulting local linear model is the dashed line.

detection systems in real-world applications, especially in safety-critical domains such as monitoring spacecraft telemetry.

Local Interpretable Model-agnostic Explanations (LIME) is a popular explanation algorithm that aims to provide interpretable explanations for the predictions made by complex machine learning models ([Ribeiro et al., 2016](#)). The LIME algorithm works by generating local, interpretable models that approximate the behaviour of the complex model around a specific instance of interest. [Fig. 6](#) presents an intuitive example of how it works. Given a complex model and a specific instance for which an explanation is desired, LIME first generates a set of perturbed versions of the instance by randomly adding noise to its feature values. It then obtains predictions from the complex model for each of these disturbed instances and uses these predictions to fit a simpler, interpretable model in the local vicinity of the instance of interest. The interpretable model is trained to predict the same outcome as the complex model for the perturbed instances while being constrained to have a simple, interpretable structure.

The resulting interpretable model can then be used to explain the behaviour of the complex model around the instance of interest. Specifically, the weights of the interpretable model can be used to

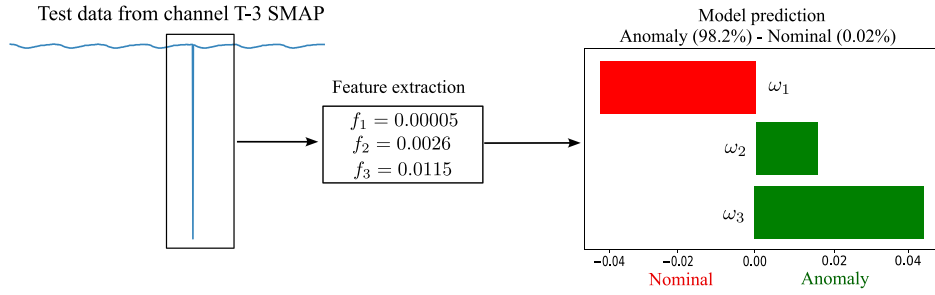


Fig. 7. LIME explanation of an instance classified as a type I anomaly in channel T-3 SMAP. CSI: 97.04, VSI: 100.0.

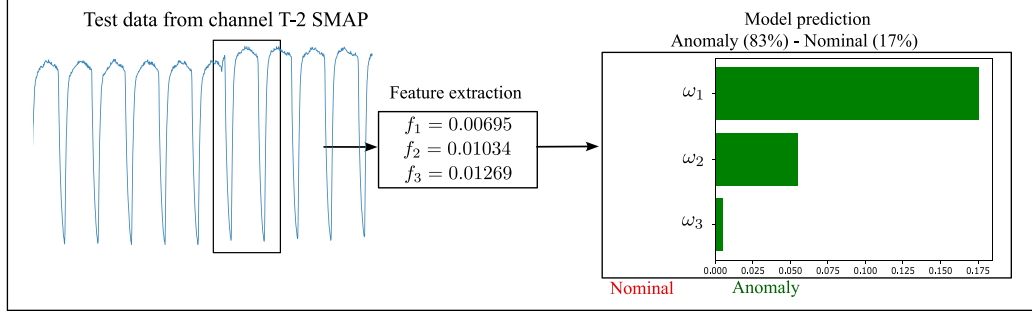


Fig. 8. LIME explanation of an instance classified as a type II anomaly in channel T-2 SMAP. CSI: 94.07, VSI: 100.0.

identify the features that are most important for the prediction made by the complex model in the local vicinity of the instance. These feature importance scores (i.e. weights) can be presented as an explanation for the prediction.

The LIME algorithm was applied to the best model obtained with Random Forest (see Table 6) to get an interpretation of the prediction on a single instance for each type of anomaly presented. Figs. 7 to 11 show the results obtained; the real test data and the window selected are marked. The feature vector of size three is presented for each instance where  $f_1$  measures the moving average prediction error,  $f_2$  is the standard deviation of STFT and  $f_3$  is the LSTM prediction error. The prediction is presented and the weights of the features  $\omega_1, \omega_2, \omega_3$  are plotted to see how much effect each feature has on the prediction of the nominal/anomaly classes. Note that the feature values must be normalized before training the complex model. Finally, the Variables Stability Index (VSI) and Coefficients Stability Index (CSI) were calculated to ensure stable explanations (Visani et al., 2022).

The explanation of an instance classified as a Type I anomaly is presented in Fig. 7. The interpretable model generated by LIME makes this prediction based on the weights of three features; the one with the highest value is  $\omega_3$  the one related to  $f_3$ , i.e. the LSTM prediction error. Is important to note that the value of  $f_2$  also affects the anomaly class, since this type of anomaly is translated into an abrupt change in frequency. On the other hand,  $f_1$ , the moving average prediction error, has more effect on the nominal class, since a change in a single value does not affect the average value of a window with  $l_w = 250$ . The opposite case occurs with an instance classified as Type II anomaly (See Fig. 8), where the main feature is  $f_1$  as expected since Type II anomalies correspond to a variation of the magnitude values. For the same reason, the frequency features  $f_2$  still affect the prediction. In addition, the effect of  $f_3$  in the prediction is not significant because the waveform of the channel does not change.

Fig. 9 presents the explanation of an instance classified as a Type III anomaly. The most important weight is  $\omega_3$ , the one associated with  $f_3$ . It is relevant to highlight that this is the only weight that contributes to the anomaly class. The main reason for this is that the only feature that captures waveform information is the LSTM prediction error. Another example is presented in Fig. 10 with an instance classified as a type IV

anomaly. In this case, the change is visible in the frequency domain, and the feature that has more impact on the classifier decision is  $f_2$ , the STFT standard deviation; the changes in magnitude, i.e.,  $f_1$  affect the prediction.

Finally, the explanation of an instance classified as a Type V anomaly is presented in Fig. 11. In this specific type, the anomaly is present by the change, not the values, and it is visible in the high effect of  $f_1$  on the prediction of the nominal class. Since an abrupt change of magnitude implies high frequency, the weight  $\omega_2$  associated with  $f_2$  has the main importance on the prediction. The same effect can be seen on the weight  $\omega_3$  related to  $f_3$  that is translated into a detected change on the wave-form of the channel resulting in an LSTM prediction error.

## 6. Conclusions and future works

This paper presents a supervised anomaly detection system for spacecraft telemetry data. The proposed approach was tested using public NASA data sets from the real SMAP and MSL missions, and outperforms previous work and other approaches from the literature.

The contributions of this paper lie in the following aspects: The proposed method is capable of detecting and treating five different types of anomalies reflected in the telemetry of spacecraft from real NASA missions, which expands the set of anomalies that are usually collected in Literature. The feature extraction phase includes information on relevant changes in magnitude, frequency and waveform, which has allowed a greater variety of anomaly types to be addressed. The classification has been implemented using three machine learning methods, which have been compared and evaluated. This has allowed us to analyse the influence of the feature extraction process on the results and opens the possibility of using other more complex methods. Finally, an explainability analysis of the proposed anomaly detection system is presented, identifying the characteristics that are most important for predicting the type of anomalies, and also providing information on the possible cause of the anomaly.

Some of the limitations of the proposed approach open the door to future work. Since noise from sensor measurements is inevitable, conducting more experiments would be an avenue to test the system's sensitivity to different types of noise and explore how different noise



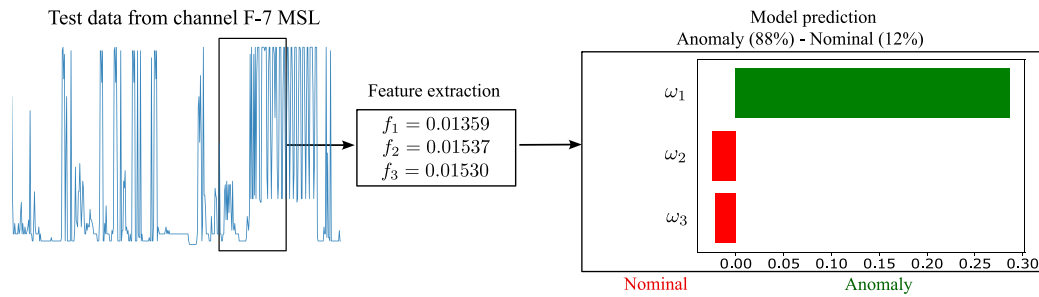


Fig. 9. LIME explanation of an instance classified as a type III anomaly in channel F-7 MSL. CSI: 99.26, VSI: 100.0.

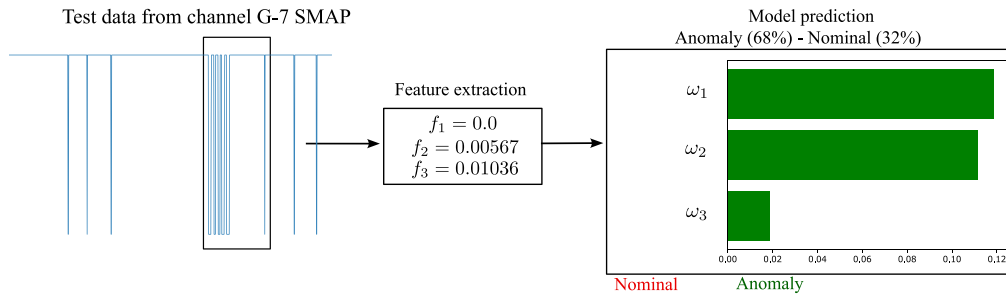


Fig. 10. LIME explanation of an instance classified as a type IV anomaly in channel G-7 SMAP. CSI: 97.78, VSI: 100.0.

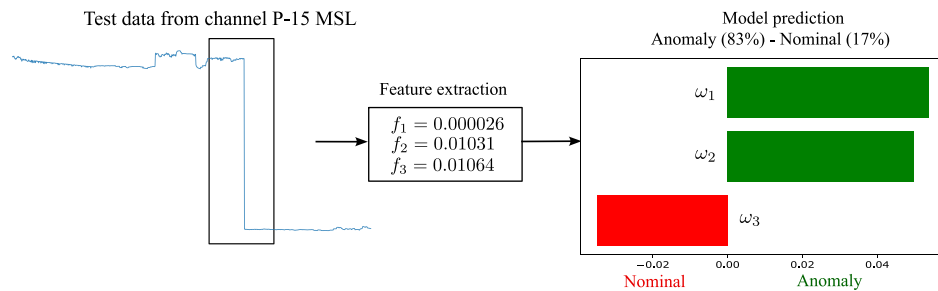


Fig. 11. LIME explanation of an instance classified as a type V anomaly in channel P-15 MSL. CSI: 100.0, VSI: 100.0.

levels affect system performance over time. As a consequence, strategies could be developed to mitigate these effects. Further research is needed to validate the effectiveness and applicability of the presented approach with data sets from different space missions to evaluate its efficiency under various conditions and contexts. With future refinements, this approach has the potential to become a valuable tool that aids the safety and success of space missions.

#### CRedit authorship contribution statement

**Sara Cuéllar:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Matilde Santos:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision. **Fernando Alonso:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing. **Ernesto Fabregas:** Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing. **Gonzalo Farias:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

This research was supported, in part, by the Chilean Research and Development Agency (ANID) under Project FONDECYT 1191188 and PhD National Grant 21221393. The Ministry of Science and Innovation of Spain under Projects PID2019-108377RB-C32 and PID2022-137680OB-C32. The Agencia Estatal de Investigación (AEI), Spain under Project PID2022-139187OB-I00. The Universidad Nacional de Educación a Distancia (UNED), Spain funding for open-access publishing.

#### References

- Abdelghafar, S., Darwish, A., Hassanien, A.E., Yahia, M., Zaghrout, A., 2019. Anomaly detection of satellite telemetry based on optimized extreme learning machine. *J. Space Safety Eng.* 6 (4), 291–298.
- Ahn, H., Jung, D., Choi, H.-L., 2020. Deep generative models-based anomaly detection for spacecraft control systems. *Sensors* 20 (7), 1991.
- Alarcon-Aquino, V., Barria, J.A., 2001. Anomaly detection in communication networks using wavelets. *IEE Proc.-Commun.* 148 (6), 355–362.
- Aminikhanghahi, S., Cook, D.J., 2017. A survey of methods for time series change point detection. *Knowl. Info. Syst.* 51 (2), 339–367.
- Baireddy, S., Desai, S.R., Mathieson, J.L., Foster, R.H., Chan, M.W., Comer, M.L., Delp, E.J., 2021. Spacecraft time-series anomaly detection using transfer learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1951–1960.

- Biswas, G., Khorasgani, H., Stanje, G., Dubey, A., Deb, S., Ghoshal, S., 2016. An application of data-driven anomaly identification to spacecraft telemetry data. In: Annual Conference of the PHM Society, Vol. 8. pp. 1–10. <http://dx.doi.org/10.36001/phmconf.2016.v8i1.2551>.
- Carlton, A., Morgan, R., Lohmeyer, W., Cahoy, K., 2018. Telemetry fault-detection algorithms: Applications for spacecraft monitoring and space environment sensing. *J. Aerosp. Inf. Syst.* 15 (5), 239–252.
- Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. *ACM Comput. Surv.* 41 (3), 1–58.
- Chen, V.C., Ling, H., 2002. Time-Frequency Transforms for Radar Imaging and Signal Analysis. Artech house.
- Chen, J., Pi, D., Wu, Z., Zhao, X., Pan, Y., Zhang, Q., 2021. Imbalanced satellite telemetry data anomaly detection model based on Bayesian LSTM. *Acta Astronaut.* 180, 232–242.
- Codetta-Raiteri, D., Portinale, L., 2014. Dynamic bayesian networks for fault detection, identification, and recovery in autonomous spacecraft. *IEEE Trans. Syst. Man Cybern. Syst.* 45 (1), 13–24.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.
- Cuéllar, S., Fariás, G., Santos, M., Alonso, F., 2022a. Preliminary results on anomaly detection and recognition in spacecraft telemetry. In: 2022 IEEE International Conference on Automation/XXV Congress of the Chilean Association of Automatic Control. ICA-ACCA, IEEE, pp. 1–6.
- Cuéllar, S., Granados, P., Fabregas, E., Curé, M., Vargas, H., Dormido-Canto, S., Fariás, G., 2022b. Deep learning exoplanets detection by combining real and synthetic data. *PLoS One* 17 (5), e0268199.
- Das, A., Rad, P., 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*.
- Dormido-Canto, S., Fariás, G., Vega, J., Dormido, R., Sánchez, J., Duro, N., Santos, M., Martín, J., Pajares, G., 2006. Search and retrieval of plasma wave forms: Structural pattern recognition approach. *Rev. Sci. Instrum.* 77 (10), 10F514.
- Du, X., Zuo, E., Chu, Z., He, Z., Yu, J., 2023. Fluctuation-based outlier detection. *Sci. Rep.* 13 (1), 2408.
- Fariás, G., Dormido-Canto, S., Vega, J., Sánchez, J., Duro, N., Dormido, R., Ochando, M., Santos, M., Pajares, G., 2006. Searching for patterns in TJ-II time evolution signals. *Fusion Eng. Des.* 81 (15–17), 1993–1997.
- Fariás, G., Fabregas, E., Dormido-Canto, S., Vega, J., Vergara, S., 2020a. Automatic recognition of anomalous patterns in discharges by applying deep learning. *Fusion Sci. Technol.* 76 (8), 925–932.
- Fariás, G., Fabregas, E., Dormido-Canto, S., Vega, J., Vergara, S., 2020b. Automatic recognition of anomalous patterns in discharges by recurrent neural networks. *Fusion Eng. Des.* 154, 111495.
- Fernández, M.M., Yue, Y., Weber, R., 2017. Telemetry anomaly detection system using machine learning to streamline mission operations. In: 2017 6th International Conference on Space Mission Challenges for Information Technology. SMC-IT, IEEE, pp. 70–75.
- Fuertes, S., Picart, G., Tournet, J.-Y., Chaari, L., Ferrari, A., Richard, C., 2016. Improving spacecraft health monitoring with automatic anomaly detection techniques. In: 14th International Conference on Space Operations. p. 2430.
- Fujimaki, R., Yairi, T., Machida, K., 2005. An approach to spacecraft anomaly detection problem using kernel feature space. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. pp. 401–410.
- Gao, J., Song, X., Wen, Q., Wang, P., Sun, L., Xu, H., 2020. Robustad: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint arXiv:2002.09545*.
- Gao, Y., Yang, T., Xing, N., Xu, M., 2012. Fault detection and diagnosis for spacecraft using principal component analysis and support vector machines. In: 2012 7th IEEE Conference on Industrial Electronics and Applications. ICIEA, IEEE, pp. 1984–1988.
- Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A., Veeramachaneni, K., 2020. Tadgean: Time series anomaly detection using generative adversarial networks. In: 2020 IEEE International Conference on Big Data. Big Data, IEEE, pp. 33–43.
- Goutte, C., Gaussier, E., 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: European Conference on Information Retrieval. Springer, pp. 345–359.
- Hansun, S., 2013. A new approach of moving average method in time series analysis. In: 2013 Conference on New Media Studies. ConMedia, IEEE, pp. 1–4.
- 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.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2017. An introduction to statistical learning: with applications in R. Springer Texts in Statistics. Springer.
- Jiang, H., Zhang, K., Wang, J., Wang, X., Huang, P., 2019. Anomaly detection and identification in satellite telemetry data based on pseudo-period. *Appl. Sci.* 10 (1), 103.
- Li, T., Comer, M., Delp, E., Desai, S.R., Mathieson, J.L., Foster, R.H., Chan, M.W., 2019. A stacked predictor and dynamic thresholding algorithm for anomaly detection in spacecraft. In: MILCOM 2019-2019 IEEE Military Communications Conference. MILCOM, IEEE, pp. 165–170.
- Li, Q., Zhou, X., Lin, P., Li, S., 2010. Anomaly detection and fault diagnosis technology of spacecraft based on telemetry-mining. In: 2010 3rd International Symposium on Systems and Control in Aeronautics and Astronautics. IEEE, pp. 233–236.
- Liu, D., Pang, J., Song, G., Xie, W., Peng, Y., Peng, X., 2017. Fragment anomaly detection with prediction and statistical analysis for satellite telemetry. *IEEE Access* 5, 19269–19281.
- Liu, L., Tian, L., Kang, Z., Wan, T., 2023. Spacecraft anomaly detection with attention temporal convolutional networks. *Neural Comput. Appl.* 1–9.
- Malhotra, P., Vig, L., Shroff, G., Agarwal, P., et al., 2015. Long short term memory networks for anomaly detection in time series. In: ESANN, Vol. 2015. p. 89.
- Martínez-Heras, J.-A., Donati, A., 2014. Enhanced telemetry monitoring with novelty detection. *AI Mag.* 35 (4), 37–46.
- Memarzadeh, M., Matthews, B., Avrek, I., 2020. Unsupervised anomaly detection in flight data using convolutional variational auto-encoder. *Aerospace* 7 (8), 115.
- Meng, H., Zhang, Y., Li, Y., Zhao, H., 2020. Spacecraft anomaly detection via transformer reconstruction error. In: Proceedings of the International Conference on Aerospace System Science and Engineering 2019. Springer, pp. 351–362.
- Murtagh, F., 1991. Multilayer perceptrons for classification and regression. *Neurocomputing* 2 (5–6), 183–197.
- Nassar, B., Hussein, W., Mokhtar, M., 2015. Space telemetry anomaly detection based on statistical PCA algorithm. *Int. J. Electron. Commun. Eng.* 9 (6), 637–645.
- Obied, M.A., Ghaleb, F.F., Hassanien, A.E., Abdelfattah, A.M., Zakaria, W., 2023. Deep clustering-based anomaly detection and health monitoring for satellite telemetry. *Big Data Cogn. Comput.* 7 (1), 39.
- OMeara, C., Schlag, L., Wickler, M., 2018. Applications of deep learning neural networks to satellite telemetry monitoring. In: 2018 Spaceops Conference. p. 2558.
- Pan, D., Song, Z., Nie, L., Wang, B., 2020. Satellite telemetry data anomaly detection using bi-lstm prediction based model. In: 2020 IEEE International Instrumentation and Measurement Technology Conference. I2MTC, IEEE, pp. 1–6.
- Pang, J., Liu, D., Peng, Y., Peng, X., 2018. Anomaly detection for satellite telemetry series with prediction interval optimization. In: 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control. SDPC, IEEE, pp. 408–414.
- Pang, J., Liu, D., Peng, Y., Peng, X., 2019. Collective anomalies detection for sensing series of spacecraft telemetry with the fusion of probability prediction and Markov chain model. *Sensors* 19 (3), 722.
- Pilastre, B., Boussouf, L., d'Escrivan, S., Tournet, J.-Y., 2020. Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning. *Signal Process.* 168, 107320.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17, 2016. pp. 1135–1144.
- Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.-R., 2021. A unifying review of deep and shallow anomaly detection. *Proc. IEEE* 109 (5), 756–795.
- Sakurada, M., Yairi, T., 2014. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. pp. 4–11.
- Schapire, R.E., 2013. Explaining adaboost. *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, 37–52.
- Song, Y., Yu, J., Tang, D., Han, D., Wang, S., 2020. Telemetry data-based spacecraft anomaly detection using generative adversarial networks. In: 2020 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence. ICSMD, IEEE, pp. 297–301.
- Tariq, S., Lee, S., Shin, Y., Lee, M.S., Jung, O., Chung, D., Woo, S.S., 2019. Detecting anomalies in space using multivariate convolutional LSTM with mixtures of probabilistic PCA. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2123–2133.
- Tartakovsky, A.G., Polunchenko, A.S., Sokolov, G., 2012. Efficient computer network anomaly detection by changepoint detection methods. *IEEE J. Sel. Top. Sign. Process.* 7 (1), 4–11.
- Thill, M., Konen, W., Bäck, T., 2017. Time series anomaly detection with discrete wavelet transforms and maximum likelihood estimation. In: Intern. Conference on Time Series. ITISE, 2, pp. 11–23.
- Visani, G., Bagli, E., Chesani, F., Poluzzi, A., Capuzzo, D., 2022. Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. *J. Oper. Res. Soc.* 73 (1), 91–101.
- Wan, P., Zhan, Y., Jiang, W., 2019. Study on the satellite telemetry data classification based on self-learning. *IEEE Access* 8, 2656–2669.
- Wang, Y., Gong, J., Zhang, J., Han, X., 2022. A deep learning anomaly detection framework for satellite telemetry with fake anomalies. *Int. J. Aerosp. Eng.* 2022, 1–9.
- Wu, J., Yao, L., Liu, B., Ding, Z., Zhang, L., 2020. Combining OC-SVMs with LSTM for detecting anomalies in telemetry data with irregular intervals. *IEEE Access* 8, 106648–106659.
- Wu, J., Zeng, W., Yan, F., 2018. Hierarchical temporal memory method for time-series-based anomaly detection. *Neurocomputing* 273, 535–546.
- Xiang, G., Lin, R., 2021. Robust anomaly detection for multivariate data of spacecraft through recurrent neural networks and extreme value theory. *IEEE Access* 9, 167447–167457.
- Xie, L., Pi, D., Zhang, X., Chen, J., Luo, Y., Yu, W., 2021. Graph neural network approach for anomaly detection. *Measurement* 180, 109546.

- Xiong, L., Ma, H.-D., Fang, H.-Z., Zou, K.-X., Yi, D.-W., 2011. Anomaly detection of spacecraft based on least squares support vector machine. In: 2011 Prognostics and System Health Management Conference. IEEE, pp. 1–6.
- Yairi, T., Inui, M., Yoshiki, A., Kawahara, Y., Takata, N., 2010. Spacecraft telemetry data monitoring by dimensionality reduction techniques. In: Proceedings of SICE Annual Conference 2010. IEEE, pp. 1230–1234.
- Yairi, T., Takeishi, N., Oda, T., Nakajima, Y., Nishimura, N., Takata, N., 2017. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction. *IEEE Trans. Aerosp. Electron. Syst.* 53 (3), 1384–1401.
- Yu, J., Song, Y., Tang, D., Han, D., Dai, J., 2021. Telemetry data-based spacecraft anomaly detection with spatial-temporal generative adversarial networks. *IEEE Trans. Instrum. Meas.* 70, 1–9.
- Yu, B., Yu, Y., Xu, J., Xiang, G., Yang, Z., 2023. MAG: A novel approach for effective anomaly detection in spacecraft telemetry data. *IEEE Trans. Ind. Inform.*.
- Zeng, Z., Jin, G., Xu, C., Chen, S., Zhang, L., 2022. Spacecraft telemetry anomaly detection based on parametric causality and double-criteria drift streaming peaks over threshold. *Appl. Sci.* 12 (4), 1803.
- Zhang, L., Yu, J., Tang, D., Han, D., Tian, L., Dai, J., 2020. Anomaly detection for spacecraft using hierarchical agglomerative clustering based on maximal information coefficient. In: 2020 15th IEEE Conference on Industrial Electronics and Applications. ICIEA, IEEE, pp. 1848–1853.
- Zheng, L., Guang, J., Tang, S.H., 2016. Fluctuation feature extraction of satellite telemetry data and on-orbit anomaly detection. In: 2016 Prognostics and System Health Management Conference (PHM-Chengdu). IEEE, pp. 1–5.