



Research Article

A Deep Learning Anomaly Detection Framework for Satellite Telemetry with Fake Anomalies

Yakun Wang ¹, Jianglei Gong,^{1,2} Jie Zhang,¹ and Xiaodong Han ¹

¹Institute of Telecommunication and Navigation Satellites, China Academy of Space Technology, Beijing, China

²School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Xiaodong Han; willingdong@163.com

Received 2 September 2021; Revised 5 November 2021; Accepted 15 December 2021; Published 19 January 2022

Academic Editor: Paolo Castaldi

Copyright © 2022 Yakun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reducing satellite failures and keeping satellites healthy in orbit are important issues. Current satellite systems have developed modules to detect anomalies on board. However, they only target a subset of anomaly types and heavily rely on expert knowledge. To address these limitations, this paper proposes a data-driven anomaly detection framework to detect point anomalies. We first propose the Deviation Divide Mean over Neighbors (DDMN) method to figure out the fake anomaly problem caused by data errors in the satellite telemetry data. Then, we use the Long Short-Term Memory (LSTM), a deep learning method, to model the multivariable time-series data, and a Gaussian model to detect anomalies. We applied our approach to the telemetry data collected from sensors on an in-orbit satellite for more than two years and demonstrate its superiority. Moreover, we explored what conditions could lead to false alarms. The approach proposed has been deployed to the ground station to monitor the health status of the in-orbit satellites.

1. Introduction

An anomaly, by definition, is an event that differs from the usual behavior of the system data. Analyzing satellite telemetry data and conducting anomaly detection have been important subjects in the development of aeronautics and astronautics. Satellite is a complex system composed of many components that are interrelated and mutually restricted. However, unlike other complex systems, the hardware monitoring of satellites is difficult, and a single failure in a component or a subsystem may be fatal to the system. Consequently, it is highly important to check the behavior of the satellite during all its lifecycle to detect the divergences as soon as possible [1].

The Out-Of-Limit (OOL) method which consists of defining a nominal range, with lower and upper thresholds, is widely used in current satellites. One or more of the parameter acquisitions exceeding an upper or lower threshold will trigger an alarm. However, modern satellites are becoming increasingly sophisticated and complex. Determining the health state of these systems using the OOL

checking method is not sufficient as the number of sensors and component interactions grow. Firstly, this method is not robust enough to detect the various types of anomalies that may occur in the telemetry data. Secondly, it requires a significant amount of domain knowledge and expertise from operators of each abnormal behavior expected, each monitoring being explicitly programmed to detect a particular anomaly signature [2–6].

In order to address the shortcomings of the current telemetry monitoring system, data-driven methods have been proposed [7–12]. The main idea behind these methods is to use the telemetry stored to create a mathematical model of the nominal behavior of the satellite, since the anomalous events in this reference telemetry are rare. Unlike OOL methods, data-driven methods generate models from data and diagnosis directly from vast newly acquired telemetry, rather than building it based on human expertise. Data-driven methods can work effectively with dozens of parameters and determine if the current satellite system or subsystem behavior is nominal or not. These data-driven methods, ideally, do not make any assumption on the parameter behavior

during an anomaly or during the nominal functioning of the satellite. The model formed by data-driven techniques is also easy to update. It can be applied to a wide variety of systems even by a person without expert knowledge of them [13, 14].

When applying the data-driven methods to an actual satellite system, however, it is mandatory to choose an appropriate statistical model for representing the normal system behavior [7, 15]. This is not a trivial work because a satellite system is very complex and its telemetry data include many aspects, such as the fake anomaly problem.

This study gives a general anomaly detection framework for satellite telemetry. We make the following three contributions. First, we propose the Deviation Divide Mean over Neighbors (DDMN) method to handle the fake anomaly problem in satellite telemetry, which can be applied to both continuous variables and discrete variables. Then, we utilize a deep learning model, the long short-term memory (LSTM) networks, to model the multitelemetries time-series data of satellite and a Gaussian model to conduct the anomaly detection. Second, we applied the proposed approach to a two-and-a-half-year dataset from an in-orbit satellite with real anomalies. Through the extensive experiments, we prove the effectiveness of our anomaly detection framework. Third, in the experiment, we discover that three types of data can result in false alarms, which are fake anomalies, unknown incidents, and sparse samples. This is helpful for understanding the nature of false alarms and provides valuable information for further studies of anomaly detection problems.

2. Approach

2.1. Fake Anomaly Detection. Satellite telemetry data is usually composed of thousands of telemetries and occasionally contain exceptionally erroneous values when transmitted back to the ground, which we call fake anomalies. Fake anomalies are not caused by system failures but by errors in data conversion or transmission. Fake anomalies are discriminated from the true anomalies and should not be the target of anomaly detection. Before using a machine learning method, it is critical to remove the fake anomalies since they will damage the learned model. Fake anomalies are common in satellite telemetry data. Hence, removing fake anomalies is not a trivial work.

Figure 1 shows two telemetry variables with fake anomalies. For example, in Figure 2(a), the values of fake anomalies are close to -30, while normal values are larger than 5. A straightforward method to remove these fake anomalies is by defining an upper bound and a lower bound. However, it does not work for fake anomalies within the range as in Figure 2(b).

Notably, the fake anomalies rarely and randomly appear in the telemetry data and are very different from their neighboring time steps. Accordingly, we propose the Deviation Divide Mean over Neighbors (DDMN) method. Specifically, for a K -dimensional collection X with T samples, we use the indicator variable $s_{t,k}$ to denote if sample t of telemetry parameter k , i.e., $x_{t,k}$, is a fake anomaly. The DDMN method is as follows:

$$s_{t,k} = \left(\text{DDM}_{t,k}^{(\text{next})} > H \right) \& \left(\text{DDM}_{t,k}^{(\text{prior})} > H \right), \quad (1)$$

where $\text{DDM}_{t,k}^{(\text{next})}$ calculates the deviation dividing mean over the next N sliding windows as in equation (2), $\text{DDM}_{t,k}^{(\text{prior})}$ calculates the deviation dividing mean over the prior M sliding windows as in equation (3), and the operator $\&$ represents AND operation.

$$\text{DDM}_{t,k}^{(\text{next})} = \left| \frac{x_{t,k} - \mu^{(\text{next})}}{\mu^{(\text{next})}} \right|, \quad (2)$$

$$\text{DDM}_{t,k}^{(\text{prior})} = \left| \frac{x_{t,k} - \mu^{(\text{prior})}}{\mu^{(\text{prior})}} \right|. \quad (3)$$

In equations (2) and (3), $\mu^{(\text{next})}$ denotes the mean of values in the next N time slices, i.e., $\mu^{(\text{next})} = (1/N) \sum_{n=1}^N x_{t+n,k}$, and $\mu^{(\text{prior})}$ denotes the mean of values in the previous M time slices, i.e., $\mu^{(\text{prior})} = (1/M) \sum_{m=1}^M x_{t-m,k}$. $x_{t,k}$ is a fake anomaly if $s_{t,k} = 1$ satisfying a condition that both $\text{DDM}_{t,k}^{(\text{prior})}$ and $\text{DDM}_{t,k}^{(\text{next})}$ are larger than a given threshold H . Otherwise, $s_{t,k} = 0$ indicates $x_{t,k}$ is normal. DDMN can be applied to both continuous variables and discrete variables. When applied to discrete variables, it is necessary to encode it into numbers such as 1.1 and 1.2. In the experiments, we empirically set $M = 8$, $N = 8$, and $H = 2$.

The algorithm of DDMN is described in Algorithm 1. Specifically, given the data collection X , calculate the indicator variable $s_{t,k}$ for each value $x_{t,k}$ based on equation (1). If $s_{t,k} = 1$, i.e., $x_{t,k}$ is a fake anomaly, remove the sample $\mathbf{x}_{t,*}$ from X and update the length L based on the new dataset $X^{(\text{new})}$. If $s_{t,k} = 0$, i.e., $x_{t,k}$ is normal, update t to move to the next sample. After one round of calculation, return the preprocessed dataset $X^{(\text{new})}$ with fake anomalies removed.

2.2. LSTM Model. A satellite system or each of its subsystems usually has a number of different operating modes and changes from one mode to another over time. However, when an exception occurs, the system will work in a failure operating mode that the model never saw before. This paper utilizes the Long Short-Term Memory (LSTM) networks, a deep learning model, to learn the frequently observed operating from the data. LSTMs have been proved to be extremely suitable for processing time-series data [10, 11, 16, 17]. Since the anomalous events in the telemetry are rare, LSTM will be trained on a nominal dataset.

Deep learning is a distributed feature learning method. The main idea is to extract the essential characteristics of data by multiple progressive training layers. LSTMs are a kind of neural network especially dealing with time-series data. LSTMs allow storage of subsequent states in different time intervals through the periodic connection of the hidden layer nodes where the parameters are shared among the different parts of the model. LSTMs can make full use of historical information and the time-dependent relationship of the modeling signal [18].

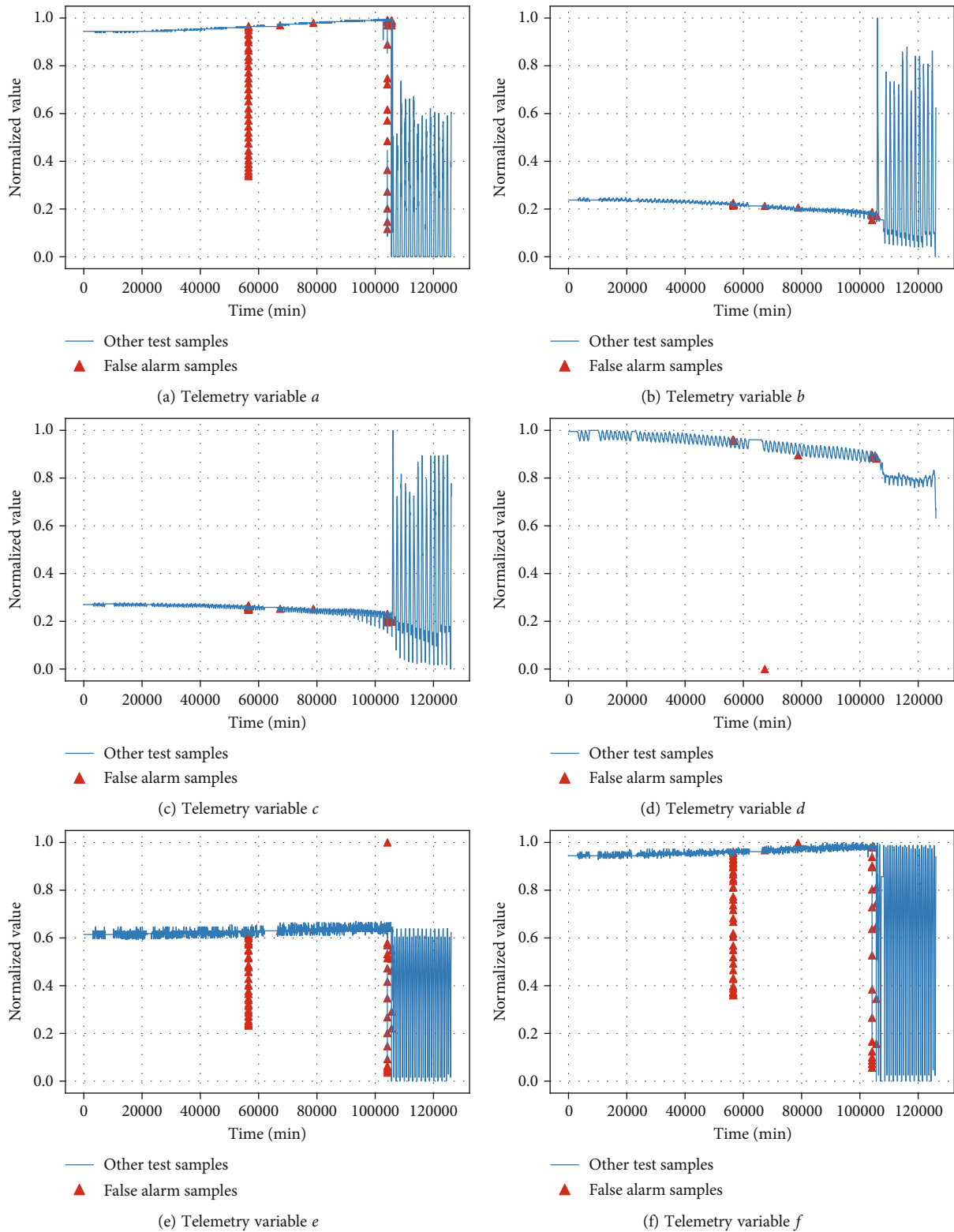


FIGURE 1: Continued.

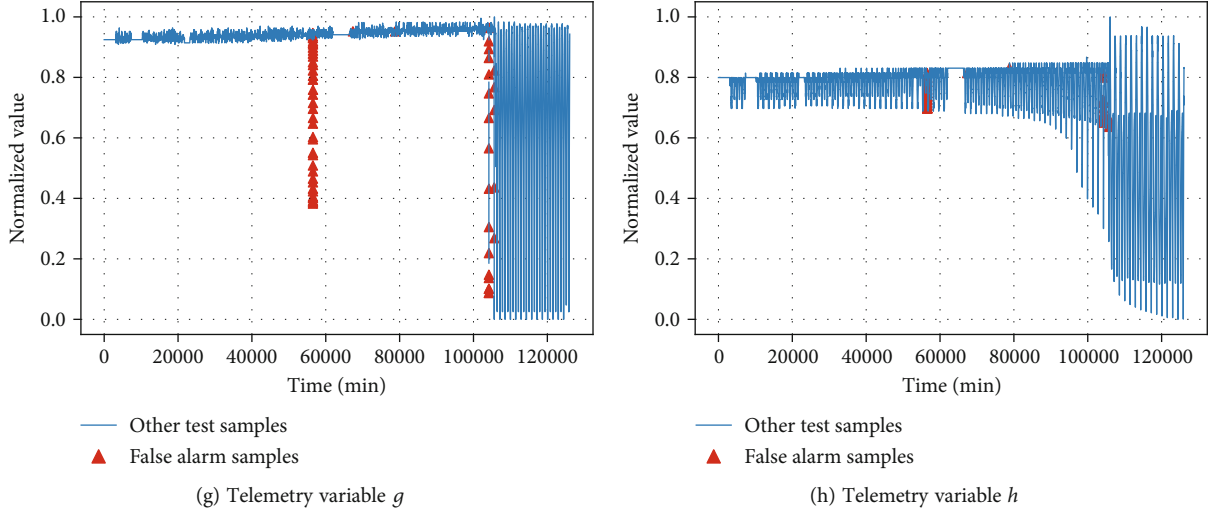


FIGURE 1: Eight telemetry variables in the testing data (red triangles are false alarm samples).

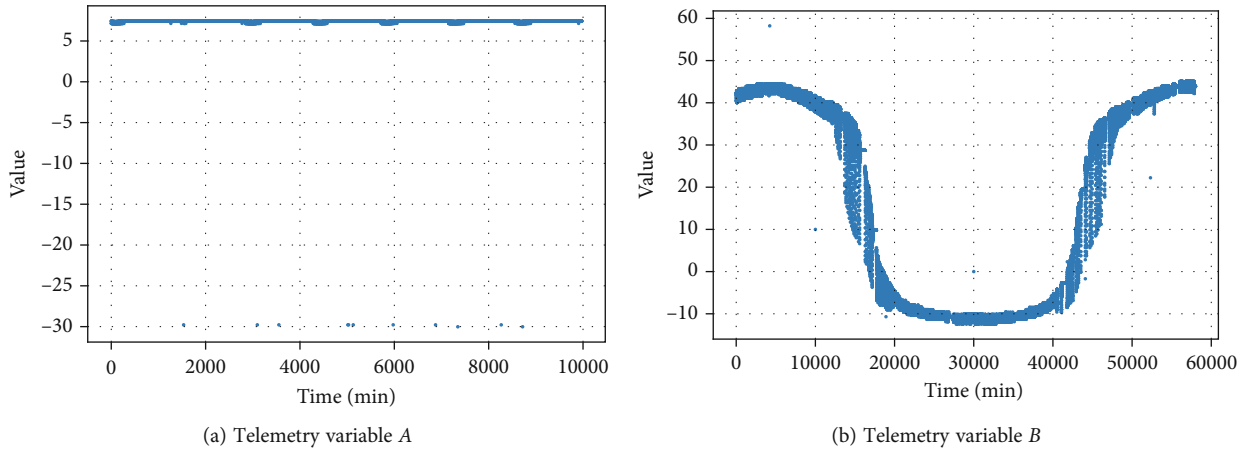


FIGURE 2: Telemetry variables with fake anomalies.

```

Input: dataset  $X$ .
Output: dataset  $X^{\text{new}}$ 
1/*Initialization */
2  $t = M$ ,  $L = |X|$ 
3 for telemetry  $k \in (1, K)$  in dataset  $X$  do
4   while  $t \in (1, T) < (L - N)$  do
5     calculate the  $\text{DDM}_{t,k}^{(\text{next})}$  using equation (2)
6     calculate the  $\text{DDM}_{t,k}^{(\text{prior})}$  using equation (3)
7     calculate the indicator variable  $s_{t,k}$  using equation (1)
8     if  $s_{t,k} = 1$  then
9       update  $X$  to  $X^{(\text{new})}$  by removing  $\mathbf{x}_{t,*}$  from  $X$ 
10       $L = |X^{(\text{new})}|$ 
11     else if  $s_{t,k} = 0$  then
12        $t = t + 1$ 
13     return  $X^{(\text{new})}$ 

```

ALGORITHM 1: Deviation Divide Mean over Neighbors (DDMN).

A typical LSTM neural network cell is configured mainly by four gates: input gate, cell gate, forget gate, and output gate. Each cell contains information at time step t on the observations that have been obtained in this step. The reading and modification of memory unit in LSTMs are realized by controlling the input gate, the forget gate, and the output gate. They are generally described by sigmoid or tanh functions [19]. Gates and cell update and output are defined as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (4)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (5)$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (6)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (7)$$

$$h_t = o_t * \tanh(c_t), \quad (8)$$

where σ is the sigmoid function, i is the input gate activation vector, f the forget gate activation function, o is the output gate activation function, and c the cell activation function.

W is the weight matrix from each cell to gate vector. h is the output of the unit. Specifically, at each time step, the LSTM unit receives input from the current state $\mathbf{x}_{t,*}$ and the hidden state h_{t-1} of the LSTMs from the previous time step. When receiving the input information, each gate will operate on data from different sources and its activation function determines whether it is active. After the forget gate f_t and the input gate i_t are transformed by the sigmoid function, they are encoded to the state of the memory cell to form a new memory cell state c_t . Finally, the memory cell state c_t forms the output h_t of the LSTM unit by the operation of the tanh function and the dynamic control of the output gate o_t [?].

In our scenario, the next time step of the time series telemetry data will be predicted based on the prior L' steps by LSTMs. After comparing the predicted value with the actual value to judge the deviation degree, the anomalies are determined, which we will discuss in the next section.

2.3. Anomaly Detection. When the model parameters have been estimated from the training data, the learned model can be utilized to generate the errors between the true value y_t and the predicted value \hat{y}_t at time t as $e_t = |y_t - \hat{y}_t|$. However, this score does not indicate whether the sample is anomalous or normal.

To set a reasonable threshold to discriminate anomalous samples from normal ones, we estimate a Gaussian model $N(\mu', \sigma')$ from the training error using maximum likelihood estimation. Then, for any value e_t , we define its anomaly score as $a_t = |(e_t - \mu')/\sigma'|$. If $a_t > H'$, x_t is considered as an anomaly. In the real application, we set the threshold $H' = 2$.

3. Experiment

3.1. Dataset and Settings. The dataset used in the experiment comes from an in-orbit satellite. Two-and-a-half-year data with more than 1.3 million samples regarding 8 telemetries are contained in the dataset. The reason we use this dataset is that a failure occurred during this period. We label the anomalies based on the occurring and end time of the failure. Finally, 0.15% of data samples are labeled as anomalous and the rest are normal. We use the early 80% data for training, which only contains the nominal data, and the remaining 20% data for testing that contains anomalies. Only the training dataset is preprocessed using the DDMN algorithm proposed in Section 2. For parameters in LSTMs, we summarize them in Table 1.

3.2. Metrics. We use the precision, recall, and F1-score to evaluate the performance of our approach. The mathematical representations of these metrics are calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (10)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (11)$$

TABLE 1: Configurations of the LSTMs.

Parameters	Configuration
Hidden layers	4
Units in hidden layers	128, 64, 32, 32
Batch size	32
Input length L'	20
Dropout	0.2
Optimizer	Adam
Loss function	Mean absolute error

where TP (True Positive) represents the number of telemetries correctly classified as anomalous, TN (True Negative) represents the number of telemetries correctly classified as normal, FP (False Positive) represents the number of telemetries incorrectly classified as anomalous, and FN (False Negative) represents the number of telemetries incorrectly classified as normal.

3.3. Fake Anomaly Detection. We qualitatively compare the result of training dataset before and after fake anomalies' removal in Figure 3. It is observed that fake anomalies are common to satellite telemetries and obscure the real structure of data. After removing fake anomalies by the proposed DDMN method, the telemetry values return to the normal range as in Figure 3(b).

For a quantitative evaluation, we compare DDMN with the following methods:

- (i) DDM^(next): calculate the deviation dividing mean of the next N time steps as in equation (2)
- (ii) DDM^(prior): calculate the deviation dividing mean of the prior M time steps as in equation (3)
- (iii) Absolute Z-SCORE [20]: calculate the absolute value of deviation dividing standard deviation of the prior M time steps as in

$$\text{Z-SCORE}_{t,j} = \left| \frac{x_{t,j} - \mu^{(\text{prior})}}{\sigma^{(\text{prior})}} \right|, \quad (12)$$

where $\mu^{(\text{prior})}$ is the mean of values in the previous M time slices, i.e., $\mu^{(\text{prior})} = (1/M) \sum_{m=1}^M x_{t-m,j}$, $\sigma^{(\text{prior})}$ is the corresponding standard deviation, i.e., $\sigma^{(\text{prior})} = \sqrt{(\sum_{m=1}^M (x_{t-m,j} - \mu^{(\text{prior})})^2) / (M - 1)}$, and $x_{t,j}$ is detected as a fake anomaly when $\text{Z-SCORE}_{t,j} > H$

- (iv) K -means [18]: K -means is a clustering method configured with two components, where we assume that only two clusters are in the dataset. One is normal, and the other one represents the fake anomaly. K -means determines the label for each sample based on its Euclidean distance to each cluster

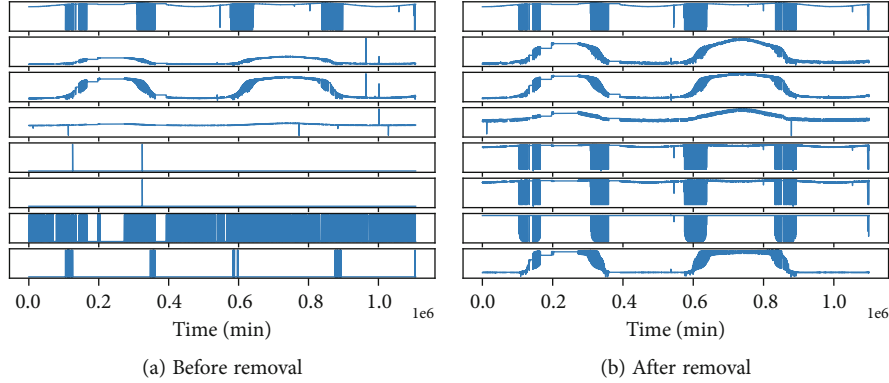


FIGURE 3: DDMN for fake anomalies' removal: (a) values of 8 telemetries over time before fake anomalies' removal; (b) values of 8 telemetries after fake anomalies' removal.

- (v) Gaussian Mixture Model (GMM) [18]: GMM is a clustering method configured with U clusters, where we assume that normal data are centered in only one cluster but fake anomalies fall into different clusters as illustrated in Figure 1. U is determined by the Bayesian information criterion [21] from the training dataset. GMM determines the label for each sample based on its generative probability from each cluster

For the ground truth, we label a certain amount of segments in each telemetry variable. As a result, 287 out of 21644 samples are labeled as fake anomalies. We calculate the average precision, recall, and F1-score over 8 telemetry variables for each method. The comparison result is shown in Table 2. First, DDMN achieves 100% precision in different settings of H . Second, DDMN always achieves a high F1-score over 90%. Third, DDMN achieves the highest F1-score of 0.969 when $H = 2$, which improves $DDM^{(next)}$, $DDM^{(prior)}$, Z-SCORE, GMM, and K-means by 93%, 92%, 30%, 61%, and 126%, respectively. Last but not the least, DDMN is robust to the variation of threshold H . In contrast, $DDM^{(next)}$, $DDM^{(prior)}$, and Z-SCORE get a low precision when H is small because they tend to detect the rapidly changing values as fake anomalies. GMM and K-means perform worst in recall because they accidentally assign fake anomalies to the normal cluster.

3.4. Anomaly Detection. In this section, we evaluate the LSTMs trained on the dataset containing fake anomalies and the dataset after fake anomalies are removed by DDMN, which we call LSTMs and LSTMs-DDMN, respectively. In the real application, we set threshold the $H' = 2$. However, for a comprehensive comparison, we compared them over different settings of H' in terms of precision, recall, and F1-score. The results are shown in Table 3. As we can see, LSTMs-DDMN outperforms LSTMs on different settings of threshold H' in both recall and F1-score. Particularly, when $H = 4$, LSTMs-DDMN improves LSTMs in terms of recall and F1-score by 49% and 27%, respectively. This is reasonable since training with the contaminated dataset has a negative effect on the learned model. In addition,

TABLE 2: DDMN and the compared methods in precision, recall, and F1-score.

(a)						
H	DDMN			$DDM^{(next)}$		
	Precision	Recall	F1-score	Precision	Recall	F1-score
2	1.0	0.939	0.969	0.34	0.947	0.501
2.5	1.0	0.936	0.967	0.654	0.946	0.773
3	1.0	0.928	0.963	0.654	0.945	0.773
3.5	1.0	0.919	0.956	0.654	0.923	0.766
4	1.0	0.9	0.951	0.654	0.907	0.76

(b)						
H	$DDM^{(prior)}$			Z-SCORE		
	Precision	Recall	F1-score	Precision	Recall	F1-score
2	0.342	0.953	0.504	0.707	0.891	0.788
2.5	0.653	0.951	0.774	0.706	0.89	0.787
3	0.653	0.946	0.773	0.814	0.837	0.825
3.5	0.653	0.924	0.765	0.814	0.837	0.825
4	0.653	0.908	0.760	0.814	0.837	0.825

(c)					
GMM			K-means		
Precision	Recall	F1-score	Precision	Recall	F1-score
0.668	0.547	0.601	0.41	0.447	0.428

LSTMs-DDMN is more robust to the variation of threshold compared to LSTMs. This is because, when H' gets larger, the F1-score of LSTMs-DDMN only drops from 0.995 to 0.904, while LSTMs drop from 0.981 to 0.713. From the perspective of entropy, the learned model of LSTMs-DDMN has a higher certainty compared to the LSTMs. In summary, first, it is not appropriate to train LSTMs with a satellite telemetry dataset containing fake anomalies. Second, the LSTMs-DDMN is effective in anomaly detection for satellite telemetry data.

TABLE 3: Comparison of LSTMs with LSTMs-DDMN in precision, recall, and F1-score.

H	Precision			Recall			F1-score	
2	0.994	0.994	0.968	0.997	0.981	0.995		
2.5	0.998	0.998	0.963	0.986	0.975	0.992		
3	0.999	0.999	0.914	0.957	0.955	0.978		
3.5	0.999	0.999	0.761	0.892	0.864	0.943		
4	0.999	0.999	0.553	0.825	0.713	0.904		
	LSTMs	LSTMs-DDMN	LSTMs	LSTMs-DDMN	LSTMs	LSTMs-DDMN		

3.5. False Alarm Study. False alarms, i.e., false positives, are normal data but predicted as anomalous. Compared with the OOL method, a weakness of the data-driven method is the false alarm problem. An anomaly detection system with too many unexpected false alarms is not suitable for the real application.

Figure 1 reports the normalized variables in the test data, where the false alarm samples are marked with red triangles generated by LSTM-DDMN. Based on the observation, we discover that three situations trigger the false alarms: (1) fake anomalies, for example, the single deviated samples in Figures 1(d) and 1(e); (2) unknown incidents such as the continuous drop in Figures 1(a), 1(e), 1(f), and 1(g); (3) sparse samples that are normal but rarely seen in the training dataset. For fake anomalies, they inevitably lead to false alarms because they are essentially some kind of anomalies. They can be avoided by engineering approaches such as a multidecision method. For unknown incidents, they may be caused by unknown events in the satellite or space. Investigating the reason can help to understand the status of the satellite. For sparse samples, they are usually generated when the satellite changes from one mode to another, which lead to infrequent samples. Some error smoothing methods were proposed to figure out this problem [11]. In summary, although the deep learning approach inevitably brings some false alarms, they are able to detect anomalies and provide valuable information to the operators about what is occurring in the system.

4. Related Work

4.1. Anomaly Detection Methods. Anomalies can be classified into the following three categories: point, contextual, and collective [18]. Point anomalies refer to individual data instances that fall within low-density regions of values. Contextual anomalies refer to anomalous data instances in a specific context. Collective anomalies refer to a collection of anomalous data instances with respect to the entire dataset. This paper mainly focuses on point anomalies.

The Out-Of-Limit (OOL) is a knowledge-driven method. However, the OOL method is not sufficient for the modern complex satellite systems with a huge number of sensors. To explore potential improvements over OOL approaches, data-driven methods are developed by learning knowledge from data using various machine learning approaches. These approaches include dimensionality reduction approaches [7, 22, 24], clustering-based approaches [8, 23–25], and nearest neighbor methods [26–28].

Deep learning, a subset of machine learning, has been widely used for satellite anomaly detection problem [10]. These approaches can be categorized into (1) prediction based and (2) reconstruction based. A typical prediction-based approach is Long Short-Term Memory (LSTM). LSTM has been proved to be effective in detecting anomalies using expert-labeled telemetry anomaly data from the Soil Moisture Active Passive (SMAP) satellite and the Mars Science Laboratory (MSL) rover, curiosity [11]. They also proposed a complementary unsupervised and nonparametric anomaly thresholding approach and false-positive mitigation strategies. The exponentially weighted average method is used to generate smoothed errors. Convolution LSTM with Mixtures of Probabilistic Principal Component Analyzers is developed for anomaly detection from Korea Multi-Purpose Satellite 2 (KOMPSAT-2) [16]. They employed both neural networks and probabilistic clustering to improve the anomaly detection performance. Statistics characteristics such as mean and standard deviation are extracted as the input of convolution LSTM. Bayesian LSTM [17] is proposed which attached the Bayesian principles to the LSTM. Then, an autoencoder is used to measure the uncertainty of samples for anomaly detection. Different approaches of LSTM based on one-to-one, many-to-one, and many-to-many network architectures are evaluated to explore the best anomaly detection structure [29]. A typical reconstruction-based approach is autoencoder [30] which compresses the input data into a lower-dimensional space and reconstructs the original data again from this representation. The reconstruction error is used to highlight the anomalous degree of data samples [17]. Autoencoder also can be used to denoise input data and extract new features which are then regarded as new features to feed into anomaly detection models [31]. Different from these works, we propose a general anomaly detection framework solving problems of fake anomaly detection and anomaly detection in satellite telemetry at the same time.

4.2. Anomaly Detection System in Aerospace Industry. NASA-Ames developed the Inductive Monitoring System (IMS) using clustering to extract a data model of the nominal telemetry data. Nominal telemetry parameter values are run through a clustering algorithm to identify nominal regions. When new telemetry data arrives, its distance to the nearest nominal region is measured, which provides a measure of a point anomaly relative to the well-defined clusters [28, 32]. Several evolutions of it have been created such as the Anomaly Monitoring Inductive Software System (AMISS) and Ames to

support the International Space Stations flight operations [4]. Nearest neighbor-based approaches have been employed by NASA's Orca tool [2] and also on board the Space Shuttle and the International Space Station [26], as well as the XMM-Newton satellite [33]. Automated Telemetry Health Monitoring System (ATHMoS), developed by the German Aerospace Center, used concepts behind the LoOP and Intrinsic Dimension Outlier Score (IDOS) outlier detection algorithms to create the Outlier Probability Via Intrinsic Dimension (OPVID) algorithm. This algorithm assigns an outlier probability that is computed via the local (continuous) Intrinsic Dimension (ID) of each point [3]. Autoencoder is also utilized to reduce the dimensions of input data. The European Space Agency (ESA) built the novelty detection system [6] that employed the density-based Local Outlier Probability (LoOP) algorithm [34] which assigns a probability to each new sample that novel telemetry behavior occurred during its respective time interval. The Japan Aerospace Exploration Agency (JAXA) developed a probabilistic clustering and dimensionality reduction method based on MPPCA [35] to model the high-dimensional, multimodal, and heterogeneous data and applied it to the telemetry data of the small demonstration satellite 4 (SDS-4) for anomaly detection [7].

5. Conclusion

This paper introduces an anomaly detection framework including a Deviation Divide Mean over Neighbors (DDMN) algorithm for fake anomaly detection and a LSTM-based anomaly detection approach. Through extensive experiments, we prove the superiority of DDMN compared with other unsupervised methods for fake anomaly detection. We also demonstrate the effectiveness of LSTMs based on DDMN for anomaly detection, i.e., LSTMs-DDMN. More than that, we discover that three types of data can result in false alarms, which are fake anomalies, unknown incidents, and sparse samples. For future work, we will collect more satellite data containing anomalies to evaluate our approach.

Data Availability

The data used in the research is from in-orbit satellites. But it has not been made public yet.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61972398 and in part by Key Research Program under Grant 2019-JCJQ-ZD-342-00.

References

- [1] C. Barreyre, L. Boussouf, B. Cabon, B. Laurent, and J.-M. Loubes, "Statistical methods for outlier detection in space telemetries," in *Space Operations: Inspiring Humankind's Future*, pp. 513–547, Springer, 2019.
- [2] S. D. Bay and M. Schwabacher, "Mining distance-based outliers in near linear time with randomization and a simple pruning rule, in," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 29–38, New York, NY, USA, 2003.
- [3] C. OMeara, L. Schlag, L. Faltenbacher, and M. Wickler, "ATHMoS: automated telemetry health monitoring system at GSOC using outlier detection and supervised machine learning," in *14th International Conference on Space Operations*, p. 2347, Daejeon, Korea, 2016.
- [4] D. Iverson, R. Martin, M. Schwabacher et al., "General purpose data-driven system monitoring for space operations," in *2009 aiaa infotech@ aerospace conference*, Seattle, WA, 2009.
- [5] S. Fuertes, G. Picart, J.-Y. Tourneret, L. Chaari, A. Ferrari, and C. Richard, "Improving spacecraft health monitoring with automatic anomaly detection techniques," in *14th International Conference on Space Operations*, p. 2430, Daejeon, Korea, 2016.
- [6] J. Martinez, "New telemetry monitoring paradigm with novelty detection, in," *Space Ops 2012*, 2012.
- [7] T. Yairi, N. Takeishi, T. Oda, Y. Nakajima, N. Nishimura, and N. Takata, "A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1384–1401, 2017.
- [8] L. Zhang, J. Yu, D. Tang, D. Han, L. Tian, and J. Dai, "Anomaly detection for spacecraft using hierarchical agglomerative clustering based on maximal information coefficient," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1848–1853, Chengdu, China, 2020.
- [9] C. Tan, G. Xu, L. Dong, H. Zhao, J. Li, and S. Zhang, "Neural network-based finite-time fault-tolerant control for spacecraft without unwinding," *International Journal of Aerospace Engineering*, vol. 2021, Article ID 9269438, 10 pages, 2021.
- [10] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: a survey," 2019, <https://arxiv.org/abs/1901.03407>.
- [11] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "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, New York, NY, USA, 2018.
- [12] S. Yan, "Understanding LSTM – a tutorial into long short-term memory recurrent neural networks," 2019, <https://arxiv.org/abs/1909.09586>.
- [13] L. Schlag, C. OMeara, and M. Wickler, "Numerical analysis of automated anomaly detection algorithms for satellite telemetry," in *2018 SpaceOps Conference*, p. 2534, Marseille, France, 2018.
- [14] K. Havelund and R. Joshi, "Experience with rule-based analysis of spacecraft logs," in *International Workshop on Formal Techniques for Safety-Critical Systems*, pp. 1–16, Springer, 2015.
- [15] V. Muthusamy and K. D. Kumar, "A novel data-driven method for fault detection and isolation of control moment gyroscopes onboard satellites," *Acta Astronautica*, vol. 180, pp. 604–621, 2021.
- [16] S. Tariq, S. Lee, Y. Shin et al., "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, New York, NY, USA, 2019.
- [17] J. Chen, D. Pi, Z. Wu, X. Zhao, Y. Pan, and Q. Zhang, “Imbalanced satellite telemetry data anomaly detection model based on Bayesian LSTM,” *Acta Astronautica*, vol. 180, pp. 232–242, 2021.
- [18] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.
- [19] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [20] D. G. Zill, *Advanced Engineering Mathematics*, Jones & Bartlett Publishers, 2020.
- [21] L. Li, R. J. Hansman, R. Palacios, and R. Welsch, “Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring,” *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 45–57, 2016.
- [22] T. Yairi, M. Nakatsugawa, K. Hori, S. Nakasuka, K. Machida, and N. Ishihama, “Adaptive limit checking for spacecraft telemetry data using regression tree learning,” in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 6, pp. 5130–5135, The Hague, Netherlands, 2004.
- [23] Y. Gao, T. Yang, M. Xu, and N. Xing, “An unsupervised anomaly detection approach for spacecraft based on normal behavior clustering,” in *2012 Fifth International Conference on Intelligent Computation Technology and Automation*, pp. 478–481, Xi’an, China, 2012.
- [24] K. Li, Y. Wu, S. Song, Y. Sun, J. Wang, and Y. Li, “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, Part G: Journal of Aerospace Engineering*, vol. 1, pp. 98–108, 2017.
- [25] T. Li and L. Chen, “Space event detection method based on cluster analysis of satellite historical orbital data,” *Acta Astronautica*, vol. 160, pp. 414–420, 2019.
- [26] D. Iverson, “Data mining applications for space mission operations system health monitoring, in,” in *SpaceOps 2008 Conference*, p. 3212, Heidelberg, Germany, 2008.
- [27] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: identifying density-based local outliers,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, vol. 29no. 2, pp. 93–104, New York, NY, USA, 2000.
- [28] S. Kannan and T. Devi, “Mining satellite telemetry data: comparison of ruleinduction and association mining techniques,” in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pp. 259–264, Crans-Montana, Switzerland, 2016.
- [29] T. A. Mahmoud, A. F. Shehab, and M. A. Elshafey, “Different long short-term memory approaches to enhance prediction-based satellite telemetry compression,” *Journal of Aerospace Information Systems*, vol. 18, no. 2, pp. 50–57, 2021.
- [30] F. Fourati and M.-S. Alouini, “Artificial intelligence for satellite communication: a review,” 2021, <https://arxiv.org/abs/2101.10899>.
- [31] C. O’Meara, L. Schlag, and M. Wickler, “Applications of deep learning neural networks to satellite telemetry monitoring,” in *2018 SpaceOps Conference*, p. 2558, Marseille, France, 2018.
- [32] D. L. Iverson, “Inductive system health monitoring,” IC-AI, 2004.
- [33] J.-A. Mart’inez-Heras and A. Donati, “Enhanced telemetry monitoring with novelty detection,” *AI Magazine*, vol. 35, no. 4, pp. 37–46, 2014.
- [34] H. P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, “LoOP: local outlier probabilities,” in *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 1649–1652, New York, NY, USA, 2009.
- [35] M. E. Tipping and C. M. Bishop, “Mixtures of probabilistic principal component analyzers,” *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.