Towards Explainable Anomaly Detection in Safety-critical Systems: Employing FRAM and SpecTRM in International Space Station Telemetry

Shota Iino[1], Hideki Nomoto[1], Takashi Fukui[2], Sayaka Ishizawa[2], Yohei Yagisawa[2], and Takayuki Hirose[1] and Yasutaka Michiura[1]

[1]*Japan Manned Space Systems Corporation, Tokyo, 100-0004, Japan*

iino.shota@jamss.co.jp
nomoto.hideki@jamss.co.jp
hirose.takayuki@jamss.co.jp
michiura.yasutaka@jamss.co.jp

[2]*JAPAN NUS Co., Ltd., Tokyo 160-0023, Japan*

fukui-t@janus.co.jp
ishizawa-syk@janus.co.jp
yagisawa-yhi@janus.co.jp

ABSTRACT

Ensuring the reliability and safety of space missions necessitates advanced anomaly detection systems capable of not only identifying deviations but also providing clear, understandable insights into their causes. This paper introduces a novel methodology for the detection of systemic anomalies in the telemetry data of the International Space Station (ISS), leveraging the synergistic application of the Functional Resonance Analysis Method (FRAM) and the Specification Tools and Requirement Methodology- Requirement Language (SpecTRM-RL). Integrated with machine learning-based normal behavior prediction model, this approach significantly enhances the explanatory of anomaly detection mechanisms. The methodology is verified and validated through its application to the thermal control system within the ISS's Japanese Experimental Module (JEM), illustrating its capacity to augment diagnostic capabilities and assist flight controllers and specialists in preserving the ISS's functional integrity. The findings underscore the importance of explainability in the machine learning-based anomaly detection of safety-critical systems and suggest a promising avenue for future explorations aimed at bolstering space system health management through improved explainability and operational resilience.

https://doi.org/10.36001/IJPHM.2024.v15i3.3857

1. INTRODUCTION

Within the orbiting International Space Station (ISS), a complex array of systems operates to maintain the necessary environmental conditions for human life and scientific experimentations. Despite the high reliability of these systems, the occurrence of anomalies remains an ever-present challenge, necessitating vigilant monitoring by flight controllers 24/7, 365 days a year. Flight controllers, organized into specialized teams such as the Fluid and Thermal Officer (FLAT), Control and Network Systems, Electrical Power, and Communication Officer (CANSEI), assess system telemetry to identify deviations that could indicate potential issues. This task, however, is complicated by the sheer volume of telemetry data and the intricate interplay of variables, making the detection of anomaly symptoms a formidable challenge. Early detection of equipment anomalies is crucial due to the extended timelines required for the manufacture, launch, and replacement of malfunctioning components. Traditional methods of monitoring single telemetry thresholds are often insufficient for early anomaly symptom detection, as many symptoms result from complex interactions among multiple variables. The limitations of current anomaly detection methodologies underscore the need for an approach that can account for the dynamic and complex nature of the ISS's operational environment. While machine learning-based methods have shown promise in identifying anomalies through the analysis of telemetry data, they often fall short in offering the explanatory depth required for flight controllers to quickly understand and react to emerging issues. In response to these challenges, this paper introduces a novel methodology for explainable symptom detection in the ISS's telemetry data. By integrating the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2012) and the Specification Tools and Requirement Methodology-Requirement Language (SpecTRM-RL) (Leveson, 2003) with machine learning techniques, our approach aims to enhance the detection of anomaly symptoms. FRAM provides a framework for understanding the causal relationships and interactions among system functions that can lead to unexpected outcomes. Meanwhile, SpecTRM-RL offers a formal method to elucidate the combinations and ranges of variables involved in anomaly detection, thereby improving the clarity and utility of the information provided to flight controllers. Together, these methodologies enable a more explainable, efficient, and effective approach to monitoring the ISS, promising significant advancements in the management of space system health and safety.

2. RELATED WORKS

The prevalent automatic anomaly detection method in spacecraft telemetry, known as the out-of-limits approach (Fuertes, 2016), triggers an alarm when telemetry values fall outside predefined normal ranges. This method's primary limitation lies in the variability of normal ranges based on operational modes and inter-telemetry relationships, complicating the setting and updating of thresholds (Chandola, 2009). To address these challenges, recent studies have shifted towards employing normal behavior models for anomaly detection, offering a more dynamic and context- sensitive solution to the limitations inherent in out-of-limit methods (Ahn, 2020; Hayton, 2007; Pilastre, 2020; Yairi, 2017; Wang, 2019). Numerous researchers have proposed automatic anomaly detection methods, with machine- learning-based approaches gaining widespread adoption for their effectiveness. In the realm of aerospace applications, Wang et al. (2019) introduced a method for diagnostic health monitoring of spacecraft in orbit, highlighting the precision of machine-learning anomaly detection techniques.

Here are three challenges associated with anomaly detection methods utilizing machine learning;

1. The selection of telemetry data for use in machine learning necessitates a profound understanding of the system and advanced analytical skills.
2. The decision-making process in black-box machine learning models lacks explainability.
3. Alerts based on the analysis results are not presented in a manner that is comprehensible to users.

When training machine learning models, if explanator variables are not appropriately selected, the analysis cannot be conducted correctly. To select appropriate telemetry data, it is necessary to understand the system's behavior. One effective approach to understand the system's behavior is interviewing experts and modeling their expert knowledge to visualize the system's characteristics.

When there are enough training data, a data-driven approach to select candidates of explanatory variable is also effective. Model methods such as STAMP (Leveson, 2012), Model Based Systems Engineering (MBSE), and Structured Analysis and Design Technique (SADT) (Ross, 1977) have been used to understand the system. These methods are accident-causal techniques having limitations such as focusing on either only machine or only human components. To address complex systems like ISS, it is essential to consider humans, machines, and their surrounding environment as a single socio-technical system and model them from a systemic perspective. FRAM is highly regarded for this purpose.

Next, anomaly detection of safety-critical systems lacks the explainability needed for operational application, particularly in ISS operations where flight operators require clear rationales for predictions to justify action. After an alert is released triggered by an anomaly detection, it is necessary to understand which input data is deviating from the normal state and how. GalaxAI, as demonstrated by Kostovska et al. (2021), offers a visualized analysis of spacecraft telemetry data, enhancing mission specialists' understanding through feature importance of machine learning-based anomaly detection. Similarly, Zeng et al. (2022) developed a framework for identifying complex telemetry relationships in spacecraft using established causal links, addressing the need for interpretability in anomaly detection. To understand the reason for detection, it is necessary to grasp the range of values each telemetry take when in a normal state and an abnormal state. Decision tree (Morgan, 1963) can be used to understand the conditions of each telemetry in a simple system. However, it is difficult to use for complex systems with numerous branching conditions. By utilizing the accumulation of mathematics and mathematical logic, it is possible to rigorously verify the correctness of information systems using formal methods, enabling the analysis of the range of values each telemetry can take. In this research, we use the SpecTRM, which allows us to represent a state that causes another state to occur as a combination of explanatory variables in a simple and understandable manner. This paper outlines our approach to anomaly symptom detection for ISS operations, detailing the development of an automated system to identify anomaly symptoms while providing supplementary information to elucidate the reasons behind these anomaly detections. This methodology aims to enhance situational awareness and facilitate informed decision-making in managing ISS system health.

3. METHODOLOGY

3.1. Overview

Our methodology for identifying and explaining symptoms in telemetry follows a structured three-step approach as shown in Figure 1.

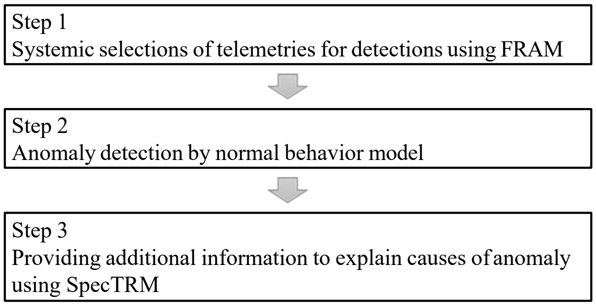

Figure 1. Flow of our proposed method

<table>
<tr><th colspan="3">Class</th><th>Aspects</th></tr>
<tr><td rowspan="5">Input</td><td colspan="2">Trigger</td><td>Input</td></tr>
<tr><td colspan="2">Prior condition</td><td>Precondition</td></tr>
<tr><td rowspan="3">Posterior condition</td><td>change output</td><td>Control</td></tr>
<tr><td rowspan="2">stop output</td><td>Resource</td></tr>
<tr><td>Time</td></tr>
<tr><td colspan="3">Output</td><td>Output</td></tr>
</table>

Table 1. Six aspects of FRAM

Firstly, we constructed a detailed system model to delineate the causal links between telemetry data, leveraging expert consultations and technical documents to accurately map out the complex interplay of system operations and their impact on telemetry signals. For the second step, we employed a machine learning-based approach for symptom detection. The final step provided a comprehensive explanation for any detected symptoms, thereby offering clear insights into the underlying issues. This process not only facilitates the early identification of potential anomalies but also enhances the understanding of complex system behaviors, crucial for maintaining operational safety and resilience.

3.2. Systemic telemetry selection for anomaly symptom detection

There are two approaches for anomaly detections: One is using many telemetries for data-driven anomaly detections. Feature extraction can be executed without expert knowledge and can apply to a large-scale system. However, it is difficult to detect anomaly symptoms for particular subsystems if there are many telemetries across subsystems. The other is using only selected telemetries with expert knowledge. It is suitable for detecting specific anomaly events of targeted subsystems. In the latter approach, telemetries are usually selected by experts, who knows the targeted system well, in advance. It is difficult to choose necessary telemetries without missing any important variables. Therefore, we propose a method to analyze and model the causal links between system functionalities and telemetries with using the FRAM modelling. FRAM allows us to delve into the complicated functional interactions within identified functions or telemetry data in the system. Initially, FRAM necessitates identifying the functions of the target system, characterized by six aspects detailed in Table 1 and visually depicted as hexagons, with each vertex representing one of the aspects. These functions interconnect through these six facets, forming a sophisticated network that models the system's complexities. Analyzing the FRAM model offers profound insights into the system's operational dynamics, illuminating how it works.

Building on Iino et al. (2022; 2023) initial use of the FRAM to model the relationships among telemetries, this paper expands the scope of FRAM modeling to encompass a broader range of system functional behaviors, incorporating additional insights gathered from extensive interviews with specialists in Safety-II approach. Safety-II approach focuses on what's going right in a system (Hollnagel, 2015). This enhanced approach allows for a deeper understanding of the complex interactions within the system, providing a more comprehensive framework for analyzing telemetry data.

3.3. Anomaly symptom detection using normal behavior models

We implemented a machine learning-based anomaly symptom detection approach utilizing Random Forest (RF) (Breiman, 2001) and Long Short-Term Memory (LSTM) (Hochreiter, 1997) to analyze selected telemetries. Recently LSTM has been widely used for anomaly detection with time-series data (Malhotra, 2015). Several researchers also applied LSTM-based methods for anomaly detections of space systems such as satellites and rovers (Hundman, 2018; Fisher, 2019). During the training phase, the model was calibrated to predict the pump inverter's temperature as objective variable based on explanatory variables, effectively modeling the telemetry relationships under normal conditions. Subsequent predictions with the trained model on testing data were conducted.

Cross-validation is a technique used to evaluate the performance of machine learning models and enhance the reliability of predictions. It is especially crucial for validating the accuracy of time series forecasting models with high reliability. Unlike the traditional approach, which splits a dataset into training and test sets, time series data require learning and validation over time, making a simple split insufficient. Cross-validation involves dividing the dataset into several periods for sequential training and validation, allowing for a more reliable evaluation of the model's performance across the entire dataset. Cross-validation was conducted to assess the prediction accuracy of the normal behavior model.

We utilized a grid search to fine-tune the hyperparameters of our RF and LSTM models, incorporating selected telemetries as explanatory variables through the model selection process. Grid search methodically explores a specified range of hyperparameters to identify the most effective settings. Following the tuning phase, we employed cross-validation to rigorously evaluate the model's accuracy, ensuring our findings were robust and reliable.

Different models for anomaly symptom detection were made based on their purpose of analysis. Then, we chose an appropriate model comparing results based on qualitative and quantitative criteria for selection. The Pugh concept selection method, proposed by Pugh (1981), was utilized for its ability to navigate these trade-offs by comparing models across multiple criteria to identify the most suitable solution. This method helps in evaluating models' engineering aspects, where the FRAM model aids in determining suitability. The Root Mean Square Error (RMSE) quantitatively assesses each model's predictive performance. Operational considerations, like the timeliness and clarity of anomaly symptom detection, are also crucial. Through this multidimensional analysis, an appropriate model for alert simulations was chosen.

3.4. Enhancing cause explanations with supplementary Information

Upon receiving alerts from anomaly symptom detections, it is imperative for flight controllers and specialists to ascertain the detection causes prior to initiating troubleshooting actions. Given that traditional machine learning-based methods for anomaly symptom detection often lack transparency, offering black-box explanations, we engaged in further analysis using SpecTRM-RL to refine the identification of potential causes. SpecTRM-RL employs a two-dimensional table mapping each parameter against time, with conditions marked as True (T) or False (F), based on predefined criteria involving average and standard deviation values. The SpecTRM analysis table was divided into two sections: "Normal condition" and "Abnormal conditions." Each section contains rows that list various parameters with associated conditions, and columns representing instances or data points with a status of True (T), False (F) or wildcard (*). Wildcard (*) means both T and F. This methodology aids in enhancing the explainability of anomaly symptom detections. Parameter conditions were specified using their average and standard deviation, as outlined in equation (1).

$$m_i - 4\sigma_i \leq x_i \leq m_i + 4\sigma_i, \qquad (1)$$

where x_i, m_i, and σ_i denote the value, mean, and standard deviation of the i-th parameter, respectively. The SpecTRM- RL algorithm then discerns patterns by identifying combinations of T, F, or * states for each parameter. By comparing these combinations before and after the onset of anomaly symptoms, flight controllers and specialists can effectively analyze telemetry trends, aiding in the assessment and understanding of system behaviors preceding anomalies.

This comparative approach enhances the predictive capabilities and situational awareness of operational teams.

4. EXPERIMENT

4.1. Experimental setup

To investigate the efficacy of our proposed anomaly symptom detection methodologies, we performed an experiment centered around a historical malfunction within Thermal Control Assembly-Low (TCA-L) in the Japanese Experimental Module (JEM) of ISS, specifically the pump failure on March 26, 2012, attributed to an overcurrent from upstream power supplies. Recognizing the complexity of this anomaly, induced by multiple factors and challenging to pinpoint through a singular parameter, our objective was to ascertain if symptom detection was feasible through the analysis of multi-variable data preceding the event. Utilizing ISS system data in 2012, particularly from the incident, we analyzed downlinked telemetry to ground stations, aiming to verify our detection approach against this real-world failure scenario. Utilizing telemetry data downloaded from JEM to ground systems from November 2011 to March 2012, we divided the data into training and testing sets as detailed in Table 2.

Model No	Period	Number of telemetry data
Train	2011/11/6-2012/1/30	3,725,381
Test (1st half)	2012/1/31-2012/3/1	2,421,955
Test (2nd half)	2012/3/1-anomaly	

Table 2. Number of telemetry data

This division was informed by trend analysis of the upstream current values of TCA-L. Specialists found the change of trend in the upstream current on around March 1, 2012. For training a machine-learning models, telemetry data during normal period should be used. Therefore, we utilized the data between November 2011 to January 2012 for training models. To simulate anomaly symptom detection, the test data includes data from periods where the anomaly occurred after March 1, 2012. Additionally, to assess how well the anomaly symptoms during normal periods are detected, data from the normal period between January 31, 2012, and March 1, 2012, is also included. To address missing values and mitigate noise, we averaged the raw telemetry data, provided at one-second intervals, over one-hour periods. The averaging interval was determined through data-driven analysis based on the prediction accuracy of the machine learning model. In our preliminary analysis, we built several predictive models for each of the data averaged over 1 minute, 10 minutes, and 1-hour intervals, respectively, and compared their mean RMSEs as shown in Table 3.

Example of the prediction of the models with 1 minute, 10 minutes, and 1 hour averaging intervals are shown in figures 2, 3, and 4 respectively. Based on Table 3, we chose the 1-hour averaging interval, which resulted in the lowest mean RMSE for this study.

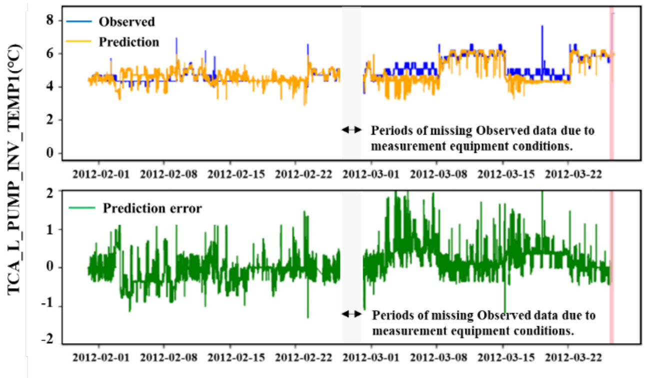

Figure 2. Example of prediction of the model using data averaged over 1-minute

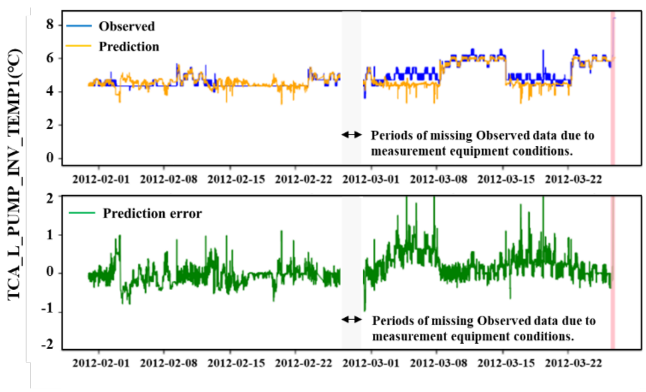

Figure 3. Example of prediction of the model using data averaged over 10-minute

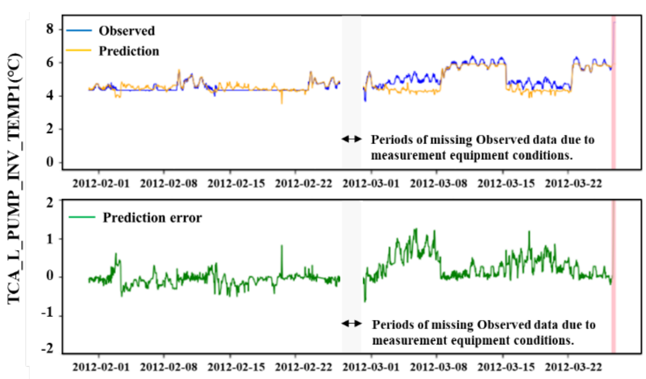

Figure 4. Example of prediction of the model using data averaged over 1-hour

Averaging Intervals	**Mean RMSE**
1-minute	0.7715714291
10-minute	0.774714286
1-hour	0.756285714

Table 3. Comparison of mean RMSE for averaging intervals

4.2. Telemetry selection through FRAM model

To select telemetries for making machine learning models, we initially understood why the system goes right. We analyzed the technical documents of the system and performed some interviews with flight controllers and specialists who know the system well. Through the analysis and interviews, we made a FRAM model to delineate the causal relationships among functions related to the anomaly in the system. We selected the temperature of the pump inverter as the objective variable as specialists identified the pump inverter system's malfunction as the probable cause of the overcurrent anomaly, guiding our focus to the pump inverter's operations within the thermal control system of JEM. Our FRAM model is represented in Figure 5, illustrating the dependencies within the pump inverter functions and the significant impact of the cabin environment on its operation.

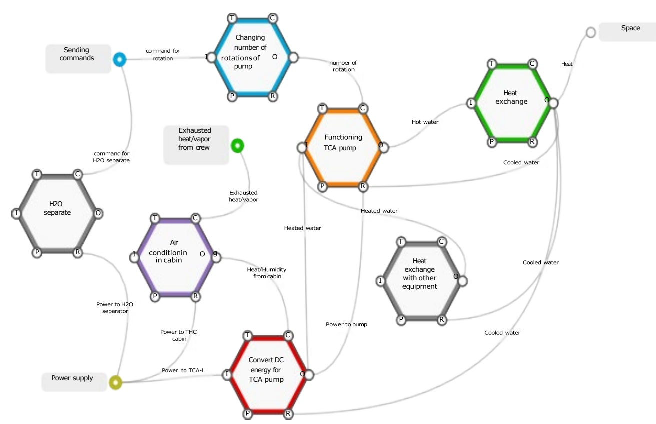

Figure 5. Systemic model through FRAM

Commands from the flight controllers control the rotational speed of the pump and the on/off state of the water separator. The pump inverter converts DC energy for TCA pump to function. The Power Supply delivers power to TCA pump while it powers the water separator. From a heat exchange perspective, heated water from other equipment flows into the TCA pump and is then expelled into space through the US module. Cooling water from the accumulator cools both the pump and the pump inverter. In addition, heat or humidity from astronauts in cabin (environment factor) affect the temperature of the TCA pump. The theory of FRAM posits that variable parameters and factors introduce system variabilities.

Furthermore, the interactions of these variabilities within the system's function network can lead to unforeseen outcomes not readily explained by traditional cause-effect relationships. This perspective underscores the complexity of system behaviors and the need for nuanced analysis in understanding and predicting system performance. Given the critical role of both constant and variable parameters in system safety, we incorporated specific telemetries for anomaly symptom detection. Our focus on the pump inverter's temperature led to selecting parameters indicative of condensation causes, notably service module partial pressure of water and cabin temperature within the JEM. The service module partial pressure of water, a key indicator of humidity levels in the ISS during 2011 and 2012, along with three additional parameters influencing pump inverter temperature, formed the basis of our analysis, highlighting the interconnectedness of system variables in predicting anomaly symptoms. Based on those characteristics of the system revealed by FRAM modelling, we made four patterns of selecting telemetries because it is important to select telemetries for training models from different views. The chosen telemetries are detailed in Table 4. For making model 1, we selected the telemetries related to TCA-L to analyze the anomaly symptoms of systems related to TCA-L such as Interface Heat Exchanger (IFHX) and TCA-L pump. For model 2, telemetries of service module partial pressure of water and cabin temp, which are related to dew point, are added besides telemetries related to TCA-L. For model 3, telemetries related to power unit of Thermal Control System (TCS) and due point are chosen to monitor the symptoms of anomaly in the trend of power. Lastly, model 4 was created with the telemetries related to cabin heat exchange, which is connected to the power unit for TCS, and due point. We compared results of anomaly symptom detections using those four models qualitatively and quantitively in next step.

4.3. Anomaly symptom detections via RF and LSTM with hyperparameter tuning

Utilizing RF regression and LSTM, we applied machine learning for anomaly symptom detection, focusing on the pump inverter's temperature influenced by explanatory variables. This modeling process, aimed at capturing normal telemetry relationships, led to predictive analyses using test data. Discrepancies between measured and predicted temperatures indicated potential anomaly symptoms, particularly in the latter test period where increased prediction errors suggested a shift towards system abnormality, culminating in a failure. Table 5 and Table 6 show hyperparameters of LSTM and RF used in the experiment.

4.4. Model selection through Pugh concept

For our quantitative analysis, we computed the RMSE for each RF model as depicted in Table 7, revealing Model 2 as the best performer with an RMSE of 0.462, and Model 4 as the second-best at 0.958.

Model No	Selected telemetries for models
Model 1 (TCA-L)	IFHX in temp IFHX out temp TCA-L LTL Control temp TCA-L pump out flow rate TCA-L pump speed
Model 2 (TCA-L + Dew point)	IFHX in temp IFHX out temp TCA-L LTL Control temp TCA-L pump out flow rate TCA-L pump speed Service module partial pressure of water Cabin temp
Model 3 (Power + Dew point)	Power Distribution Unit for TCS current out Service module partial pressure of water Cabin temp
Model 4 (Cabin heat exchange + Dew point)	Service module partial pressure of water Cabin temp Condense out pressure of water separator Cabin heat exchanger coolant out temperature Cabin heat exchanger flow rate

Table 4. Selected telemetries for each model

Parameter	Setting
Sequence length	7 hours
Number of units	600
Number of epochs	300
Batch size	50
Activation function	Relu
Optimizer	Adam
Weight initializer	Glorot uniform
Dropout	10%

Table 5. Hyperparameters of LSTM

Parameter	Setting
Number of trees	1000
Max features	1

Table 6. Hyperparameters of RF

Figure 6 shows the observed and predicted values for the objective variable across models, with blue, red, pink, and yellow lines representing the predictions from Models 1 through 4, respectively, and a grey line for the observed values. To identify abnormal symptoms, we scrutinized the deviations between predicted and observed values, considering significant deviations as indicators of abnormality. Following our comparison of models using the Pugh Concept Selection, depicted in Table 7, we found that Model 2 had the lowest RMSE, indicating high accuracy.

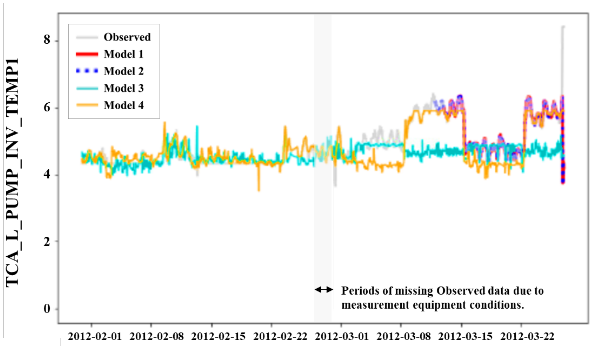

Figure 6. Observed and predicted temperatures for each model

Method	Model 1	Model 2	Model 3	Model 4
RMSE	1.307	0.462	1.095	0.958
Beginning of Symptom	1week~ 2weeks before	2days~ 3days before	2week~ 3weeks before	2week~ 3weeks before
Levels of symptoms	Low	Low	High	High

Table 7. Comparison results through Pugh Concept selection

However, Models 3 and 4 excelled in early anomaly symptom detection. Those models could detect symptoms of anomaly earlier by seeing the difference between the observed and expected values. If the differences are high, it is indicated that the parameters have clearer unusual trends. After discussions with specialists and flight controllers, we opted Model 4 for our alert simulations. Early high levels of symptoms with relatively high accuracy in RMSE provides the resilience of operations because specialists and flight controllers will have more time to plan and prepare for troubleshooting of the anomaly. This decision was driven by the paramount importance of early anomaly symptom detection, combined with the model's superior predictive accuracy, underscoring our commitment to enhancing system reliability and safety.

4.5. Comparison results of RF and LSTM through cross-validation

We divided the dataset into five periods from November 6, 2011, to March 1, 2012, for analyzing prediction error trends over time. Table 8 provides details of these intervals, enabling an examination of prediction errors' progression during the defined period. During the dip between 2012-02-22 and 2012-03-01, maintenance of the system was performed. Therefore, we excluded the data during the time.

CV	Period
1	2011/11/6 00:00:00~2011/11/28 2:00:00
2	2011/11/28 3:00:00~2011/12/20 3:00:00
3	2011/12/20 4:00:00~2012/1/15 2:00:00
4	2012/1/15 3:00:00~2012/2/6 6:00:00
5	2012/2/6 7:00:00~2012/3/1 23:00:00

Table 8. Cross-validation periods

To refine the analysis results for the RF and LSTM model of model 4, we divided telemetry data into two sets for the normal condition period, designated for training and testing to fine-tune the LSTM's hyperparameters, as shown in Table 8. We applied a one-hour averaging to each telemetry signal to manage missing data and mitigate noise issues. Following this preprocessing, we executed a grid search to optimize the hyperparameters, aiming for enhanced model performance and accuracy. We assessed the predictive accuracy of a finely tuned LSTM model through cross-validation and compared it with that of an RF model using the same method. Table 9 presents their performance across different periods, with average RMSE scores of 0.213 for LSTM and 0.174 for RF, indicating RF's superior accuracy.

CV	LSTM	RF
1	0.210	0.102
2	0.283	0.223
3	0.132	0.144
4	0.169	0.150
5	0.271	0.252
Average	0.213	0.174

Table 9. RMSE of LSTM and RF at each period

During a particular cross-validation phase marked by lower predictive accuracy, we observed significant fluctuations in the objective variable, with marked deviations from expected values, as illustrated in Figure 7.

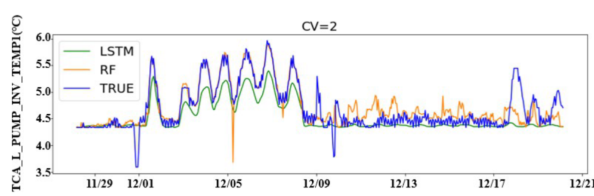

Figure 7. Observed temperatures and temperatures predicted by RF

To enhance predictive accuracy in phases of significant objective value fluctuation, we adjusted the LSTM's hyperparameters, specifically targeting high accuracy for the second cross-validation period.

This period was used as validation data, with other periods for training, employing a grid search for optimization. Subsequent cross-validation showed improved average RMSE scores: 0.136 for LSTM and 0.148 for RF, with LSTM outperforming RF, as detailed in Table 10.

CV	LSTM	RF
1	0.157	0.103
2	0.182	0.231
3	0.089	0.123
4	0.159	0.151
5	0.095	0.133
Average	0.136	0.148

Table 10. RMSE of tuned LSTM and RF at each period

4.6. Alert simulation

We conducted an alert simulation using pre-defined thresholds and compared outcomes across two, three, and four-sigma levels using RF model 4. The four-sigma threshold was selected due to its optimal alert balance discussing about the tradeoff of frequency and reliability with flight controllers and specialists. If we have less than three sigma levels, there are too many alerts based on the interviews of flight controllers. Figure 8 shows red points marking values that surpass the threshold, indicating potential alerts.

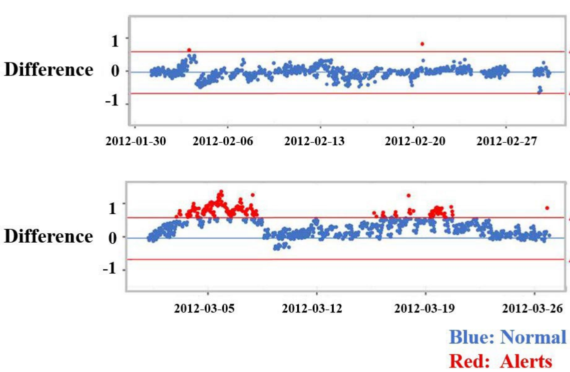

Figure 8. Alert simulations with the results of RF

These simulation findings enable the issuance of alerts to flight controllers, enhancing monitoring and response capabilities. Setting an anomaly symptom detection threshold at plus or minus four standard deviations (±4σ) from the mean reveals a striking pattern of alerts, as detailed in Table 11. Using the RF method, this approach resulted in 2 alerts from 680 data points in the first half of the year, with a significant jump to 128 alerts from 595 data points in the latter half. This substantial increase in alerts underscores the system's variable performance and the growing frequency of anomalies as the year unfolded.

	Standard deviation	Normal value range	Alarms 1st half	Alarms
RF	±4σ	-0.697～ 0.576	2/680	128/595

Table 11. Alert simulations with the results of RF

4.7. Enhancing explanation with additional information via SpecTRM

Figure 9 shows our results of SpecTRM analysis. At first, we performed a comprehensive statistical review of each telemetry data such as Service module and cabin temperature, establishing a four-sigma (4σ) threshold for each condition. Using the threshold, we obtained the conditions for each parameter. For example, service module PPH2O<11 and cabin temperature<22.

Normal condition							
Service module PPH2O < 11	T	T	F	T	T	T	*
Cabin temperature< 22	F	T	*	*	T	F	F
Condense out pressure of water separator < 2	T	T	T	T	T	T	T
Temperature of pump inverter is between 3 & 4.5	T	F	F	*	*	F	T
Cabin heat exchanger coolant out temperature < 12	T	T	F	T	F	*	F
Cabin heat exchanger flow rate< 280	T	T	T	T	T	T	T
Number of data	21	23	25	109	248	332	389

Abnormal conditions		
Service module PPH2O < 11	F	T
Cabin temperature< 22	F	F
Condense out pressure of water separator < 2	T	T
Temperature of pump inverter is between 3 & 4.5	F	F
Cabin heat exchanger coolant out temperature < 12	F	F
Cabin heat exchanger flow rate< 280	T	T
Number of data	10	118

Figure 9. Notable results of SpecTRM-RL analysis

Then, we separated the data into two group: normal and abnormal conditions. Normal conditions are data of blue points and abnormal conditions are one of red points in Figure 5 separately. Each section contains rows that list various parameters with associated conditions, and columns representing instances or data points with a status of True (T), False (F), or wildcard (*). If each parameter meets the condition, we set T. If it does not meet the condition, we put F. Using our tool for SpecTRM analysis enables us to show the complete combinations of T or F for each parameter in normal and abnormal conditions automatically. There were 21 points, which have the conditions of cabin temperature (<22) is F and other parameters' conditions are T.

The notable results reveal distinct patterns from RF analysis results: under normal operating conditions, seven combinations of telemetry states were observed, whereas under abnormal conditions, only two combinations were noted.

Notably, three parameters consistently registered false values during these abnormal conditions: cabin temperature, pump inverter temperature, and cabin heat exchanger coolant outlet temperature, indicating significant deviations from expected operational parameters.

5. DISCUSSION

5.1. Systematic telemetry selection using FRAM

FRAM modeling offers a systemic approach to understand complex systems, emphasize the holistic analysis of interactions and variabilities within the socio-technical systems. This method excels in delineating causal relationships and dependencies among system functions, particularly highlighting the interplay between constant and variable parameters. This systemic perspective is crucial for predicting unforeseen outcomes that arise from the complex interactions within the system, moving beyond traditional linear cause-and-effect analyses.

FRAM's comprehensive approach facilitates a deeper understanding of system behaviors, enhancing the ability to mitigate potential anomalies through a nuanced analysis of the interconnectedness of system variables. Although FRAM is an effective tool to develop a model of target systems, it still involves some limitations: the first limitation is that the method is time consuming to build appropriate models through manual works such as interviews with experts; the second limitation is that FRAM is essentially a qualitative method, and it is almost impossible to eliminate subjectivity in the analysis; the third limitation is that the resolution of FRAM model depends on the purpose of analysis. These limitations consistently require great workloads by analysts, and more effective ways of analysis are therefore expected. To overcome the problem, for instance, it is expected that functions can be extracted automatically from interviews of experts using Natural language processing (NLP) methods such as Generative Pre-trained Transformer (Radford, 2018) in the future. Another approach is to implement FRAM as numerical simulation models: Patriarca (2017) adopted the concept of monte-carlo simulation to evaluation of states in each function; Hirose and Sawaragi (2020) developed a simulation model based on fuzzy inference and suggested possibilities of FRAM to represent complex behaviors of target systems. Those approach is expected to reduce workloads and contribute to increasing objectivities of analysis.

5.2. Detecting anomaly symptoms using various machine learning-based methods

In the initial cross-validation, the RF model recorded a superior average RMSE score (0.174 vs. 0.213) compared to the LSTM model, indicating higher predictive accuracy for the RF model. This outcome suggests that the RF model may be more robust, particularly in handling noisy data and missing data, for this specific dataset and problem setting.

The data preprocessing steps, such as averaging every hour, and the adjustment of hyperparameters for the LSTM model, played a crucial role in enhancing model performance. Notably, during the cross-validation phase where significant fluctuations in the objective variable were observed, adjusting the LSTM's hyperparameters improved the predictive accuracy. Specifically, this suggests that recursive neural networks like LSTM are effective, especially in scenarios involving time-series data or significant fluctuations in the objective variable. It is also possible to perform anomaly prediction by combining other advanced machine learning methods with the approach proposed in this paper.

5.3. Analyzing the rationale for anomaly alert triggers

During the normal operational period, varied conditions were observed; however, unique characteristics emerged in abnormal conditions, specifically relating to cabin temperature and cabin heat exchanger coolant outlet temperature, both critical for dew point considerations. These conditions have implications for pump inverter condensation.

A Japanese astronaut on the ISS previously conducted troubleshooting tasks for this anomaly and identified a pump inverter failure caused by an overcurrent of power (JAXA, 2013). Our analysis results align with this observation. This incident underscores the importance of detailed anomaly symptom detection and rationale understanding for safety-critical systems, suggesting the need for further discussions on explanatory levels with flight controllers and specialists for future preparedness. In alerting flight controllers, it's crucial to include information on potential causes behind detected symptoms. SpecTRM-RL's identification of condition combinations further aids in monitoring telemetry trends, pinpointing water vapor pressure and cabin temperature in JEM as critical factors influencing condensation, with statistical analysis affirming these findings.

Although SpecTRM-RL is an effective tool to analyze the combinations of parameters, it still has a limitation: it takes lots of time to complete the analysis due to high number of combinations of parameters if there are many telemetries are used for making models. To overcome the problem, we can use SpecTRM analysis by just comparing the trends of telemetries between normal and abnormal period. The method will reveal the trend change of some combinations of telemetries. It will help us to find the rationale of anomaly symptom detections.

5.4. Future applications

In discussing future applications, Wyatt et al. (2000) emphasized the importance of monitoring of nominal conditions for the purposes of generating reports, adaptive alarm limits and model-based reasoning for future planetary exploration mission. Deep space missions require autonomous operations to cope with long communication delays. In addition, Prognostics and Health Management for remote crew health maintenance are demanded for future missions.

Predictive diagnostics to provide early and actionable real-time warnings for crews will be required (Popov, 2019). Additionally, the need for Prognostics and Health Management systems is emphasized for remote crew health maintenance, with predictive diagnostics playing a key role in providing early and actionable warnings. Our proposed method, offering interpretable assessments and warnings, is poised to significantly enhance health monitoring and symptom detection in space systems, contributing to the safety and operational resilience of future missions.

6. CONCLUSION

In our conclusion, we introduced a novel methodology for explainable symptom detection in telemetry, integrating the Functional Resonance Analysis Method (FRAM) and Specification Tools and Requirement Methodology-Requirement Language (SpecTRM-RL) with machine learning techniques. This approach facilitates systemic analysis, addressing the challenges of complex, multifactorial interactions that previous studies struggled with. Our method was verified and validated through an experiment using 2012 telemetry data from the thermal control system of Japanese Experiment Module (JEM), successfully identifying anomaly symptoms alongside their potential causes. Future enhancements will focus on refining condition thresholds for each parameter and alert mechanisms to improve detection accuracy and timeliness. Further validation across different systems and interface improvements are essential for practical application.

ACKNOWLEDGEMENT

The authors thank some JAXA Flight controllers and specialists of Japanese Experiment Module (JEM) of International Space Station for the interviewing.

REFERENCES

Ahn, H., Jung, D., & Choi, H. L. (2020). Deep generative models-based anomaly detection for spacecraft control systems. *Sensors*, 20(7), 1991.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3), 1-58.

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res*., 20(177), (p. 1-81).

Fuertes, S., Picart, G., Tourneret, 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.* 16-20 May 2016, Korea.

Hayton, P., Utete, S., King, D., King, S., Anuzis, P., & Tarassenko, L. (2007). Static and dynamic novelty detection methods for jet engine health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 493-514.

Hirose, T., & Sawaragi, T. (2020). Extended FRAM model based on cellular automaton to clarify complexity of socio-technical systems and improve their safety. *Safety science*, 123, 104556.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), (p.1735-1780).

Hollnagel, E. (2012). FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems. *Crc Press*.

Hollnagel, E., Wears, R. L., & Braithwaite, J. (2015). From Safety-I to Safety-II: a white paper. The resilient health care net: the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia.

Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. August 19–23, 2018, London, United Kingdom. doi:10.1145/3219819.3219845

Iino, S., Nomoto, H., T., Sasaki, M., Ishizawa, Y., Fukui, S, Michiura, Y., Hirose, ... & Wada, M. (2022). Explainable symptom detection in telemetry of ISS with Random Forest and SpecTRM. *Proceedings of 2022 IEEE Aerospace Conference*. 5-12 March, Big Sky, MO. doi: 10.1109/AERO53065.2022.9843739

Iino, S., Nomoto, Fukui, T., Yohei, Y., Ishizawa, Y., H., Hirose, T., Michiura, Y., ... & Shibayama, H. (2023). Time-series anomaly detection in telemetry of ISS providing the reasons with FRAM and SpecTRM. *Proceedings of 2023 IEEE Aerospace Conference*. 4-11 March, Big Sky, MO. doi: 10.1109/AERO55745.2023.10115985

Japan Aerospace Exploration Agency (JAXA). (2013). Japanese Manned Space Technology brought by Development and Operation of International Space Station (ISS), Japanese Experiment Module KIBO. *JAXA-SP-12-015*, 1 – 227.

Kostovska, A., Petković, M., Stepišnik, T., Lucas, L., Finn, T., Martínez-Heras, J., & Kocev, D. (2021). GalaxAI: Machine learning toolbox for interpretable analysis of spacecraft telemetry data. *Proceedings of 2021 IEEE 8th International Conference on Space Mission Challenges for Information Technology (SMC-IT).* 26-30 July 2021. Pasadena, CA. doi: 10.1109/SMC-IT51442.2021.00013

Leveson, N. G., Reese, J. D., & Heimdahl, M. P. (2003). SpecTRM: A CAD system for digital automation. *Proceedings of Digital Avionics Systems Conference.* 31 October 1998 - 07 November 1998. Bellevue, WA. doi: 10.1109/DASC.1998.741474

Leveson, N. (2012). Engineering a Safer World: Systems Thinking Applied to Safety (Engineering Systems). *The MIT Press*.

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.* 22-24 April 2015, Bruges Belgium. ISBN 978-287587014-8

Morgan, J.N., Sonquest, J.A. (1963). Problems in the analysis of survey data and a proposal. *J. Amer.* Statist. Assoc. 58, 415–435.

Patriarca, R., Di Gravio, G., & Costantino, F. (2017). A Monte Carlo evolution of the Functional Resonance Analysis Method (FRAM) to assess performance variability in complex systems. *Safety science*, 91, 49-60.

Pilastre, B., Boussouf, L., d'Escrivan, S., & Tourneret, J. Y. (2020). Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning. *Signal Processing*, 168, 107320.

Popov, A., Fink, W., Hess, A., & Tarbell, M. A. (2019). A paradigm shift from telemedicine to autonomous human health and performance for long-duration space missions. *International Journal of Prognostics and Health Management*, 10, 001.

Pugh, S. (1981, March). Concept selection: a method that works. *Proceedings of the International conference on Engineering Design* (p. 497-506).

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *Preprint*.

Ross, D. (1977). Structured Analysis: A Language for Communicating Ideas. *IEEE Transactions on Software Engineering,* 3(1), Special Issue on Requirements Analysis, 16-34.

Wang, C., Lu, N., Cheng, Y., & Jiang, B. (2019). A telemetry data based diagnostic health monitoring strategy for in- orbit spacecrafts with component degradation. *Advances in Mechanical Engineering*, 11(4), 1687814019839599.

Wyatt, E., & Nilsen, E. (2000). Flight System Autonomy Needs for Planetary Exploration Missions.

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 Transactions on Aerospace and Electronic Systems*, 53(3), 1384-1401.

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. *Applied Sciences*, 12(4), 1803.

BIOGRAPHIES

Shota Iino earned his B.S. in physics and mathematics from the University of Sydney in 2008, followed by a Master of System Engineering from Keio University and a Ph.D. from the University of Tsukuba. His research spans safety analysis of space systems and machine learning applications across various sectors at Japan Manned Space Systems (JAMSS) and the Center for Cybernics Research at the University of Tsukuba. He has also contributed as a flight controller for the Japanese Experiment Module (KIBO) within the International Space Station Program and the H-II Transfer Vehicle (HTV).

Hideki Nomoto initiated IV&V (Independent Verification & Validation) for the International Space Station Program in 1996. From 2005 to 2006, he worked at Aero-Astro Lab of MIT as the invited researcher. In 2014 he initiated Resilience Engineering at Japan Aerospace Exploration Agency (JAXA). At JAXA, he built the new generation vehicle (HTV-X) architecture using the Resilience Engineering with collaboration with Prof. Erik Hollnagel. He received Ph.D from University of Japan in 2018. He has been the manager of IV&V Research Lab at Japan Manned Space Systems (JAMSS).

Takashi Fukui received a Master of Science from Hokkaido University in 2006. After he joined Japan NUS Co., Ltd., an energy and environment consulting firm, he has been engaged in data analysis and AI system development in various disciplines, including anomaly detection for space systems and industrial plants.

Yohei Yagisawa received a Master of Environment from University of Tokyo in 2017. After he joined a consulting firm, JAPAN NUS Co., Ltd., he has performed data analysis of various fields, including chemical plants data analysis to develop AI prediction system, assessment of the impact of climate change and anomaly detection of space system.

Sayaka Ishizawa received a Master of Environmental Science from Toho University in 2011. After she joined a consulting firm, JAPAN NUS Co., Ltd., where she works as a data scientist to verify the effectiveness of countermeasures and propose improvements. She has worked on detecting signs of anomalies in space equipment systems, cultivation diagnosis using satellite data in the agricultural field, and evaluating the impact of climate change using river simulations.

Takayuki Hirose received his Ph.D. from Kyoto University in 2020. He has been working on the research of Resilience Engineering and specifically the development of a simulation model based on FRAM. He is currently a researcher of IV&V Research Lab at Japan Manned Space Systems (JAMSS) and engaged in the same issues in order to apply the knowledge of Resilience Engineering to various fields, beyond the space industry.

Yasutaka Michiura received a B.S. in physics from Toho University, Japan, in 2006. He is currently working on the safety and security analysis of software onboard spacecraft at Japan Manned Space Systems Corporation (JAMSS) and engaged in research on safety analysis technology for automatic control systems such as artificial satellites and automobiles using FRAM.