The Journal of Systems & Software 204 (2023) 111768

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Adaptive robustness evaluation for complex system prognostics and health management software platform☆

He Liu [a,b], Cheng Wei [a,*], Bo Sun [b,*], Yinxue Zeng [c]

[a] School of Astronautics, Harbin Institute of Technology, Harbin, 150080, China
[b] Beijing Institute of Spacecraft System Engineering, China Academy of Space Technology, Beijing, 100086, China
[c] College of Electrical Engineering, Sichuan University, Chengdu, 610065, China

ARTICLE INFO

Article history:
Received 27 January 2023
Received in revised form 15 May 2023
Accepted 25 May 2023
Available online 1 June 2023

Keywords:
Robustness evaluation
PHM software platform
Machine learning
Complex system

ABSTRACT

With the quantity and complexity of various complex systems have increased, their safety and reliability face many challenges. The prognostics and health management (PHM) software platform for complex systems is one of the important tools to ensure the reliability and safety of them. On the other hand, due to the uncertainty of sensing data, the robustness of PHM software platforms needs to be evaluated to ensure their performance. However, it is highly subjective and time-consuming for the existing robustness assessment methods which rely on expert experience. To solve these problems, a novel adaptive robustness evaluation method with Genetic Algorithm (GA) and Random Forest Algorithm (RFA) for the complex system PHM software platform is proposed. Firstly, the basic robustness indicators are extracted based on the classical metrics of PHM software platforms. Secondly, more sensitive indicators to the robustness of PHM software platforms are selected by GA from basic robustness indicators. Then, the classification is conducted by the selected indicators through RFA. Finally, the robustness of PHM software platforms is evaluated by the classification accuracy. Experiments with simulation data show that the proposed method has better performance, which is suitable for the robustness evaluation on the PHM software platform of complex systems.

1. Introduction

As the increasing number and rising complexity of various complex systems, the reliability and safety of complex systems are becoming important for their high-performance operation (Soualhi et al., 2022; Li and Ru, 2019). However, with the large amount of sensing data involved, it is essential to evaluate their robustness against various sources of uncertainty, such as data noise or system abnormal behaviors. Moreover, due to the uncertainty of the operation environment as well as the complex degradation mechanism of complex systems, there will be some inevitable anomalies or faults for them (Jiang et al., 2021). On the other hand, the complex system is generally difficult to develop, produce, operate and maintain. If the anomalies or faults of complex systems do not be detected and handled immediately, it will cause catastrophic accidents and large economic losses (Ji et al., 2019; Wong et al., 2017, 2010). To improve the safety and reliability of the normal operation of complex systems, it is significant to conduct the research of prognostics and health management (PHM) software platform (Yang et al., 2021).

Generally, the prognostics and health management (PHM) software platform of the complex system mainly includes condition monitoring, anomaly detection, fault diagnosis, remaining useful life prediction and etc. (Zhang et al., 2020; Vogl et al., 2019). The categories of their methodologies are similar to each other, which mainly include threshold-based, rule-based, model-based and data-driven methods currently (Ibrahim et al., 2019). The threshold-based method implements PHM by comparing the reference table of the normal sensing data with the real-time sensing data (Fernandez et al., 2017). However, as the number of components increases, this method is very inefficient and time-consuming. The second method mainly relies on expert knowledge to carry out PHM. Due to the large amount of expert knowledge and the rule libraries to be needed updated timely, new types of anomalies or faults cannot be detected effectively (Yairi et al., 2014). Model-based methods have difficulty in considering every possible anomaly or fault pattern of the sensing data. And it also requires expert experience to summarize and analyze the physical mechanism of the research subjects, so the established model has exact limitations such as the requirements of the extensive expert knowledge (Hundman et al., 2018).

☆ Editor: Prof. W. Eric Wong.

* Corresponding authors.
E-mail addresses: weicheng@hit.edu.cn (C. Wei), Sunboo2002@126.com (B. Sun).

https://doi.org/10.1016/j.jss.2023.111768
0164-1212/© 2023 Published by Elsevier Inc.

Compared with above-mentioned methods, the data-driven methods can implement PHM without physical models and has the advantage of high performance (Lyu et al., 2023, 2022). For example, to detect possible anomalies of complex systems, many data-driven methods have been proposed. Fuertes et al. carry out a study on anomaly detection based on the One-Class Support Vector Machine (OC-SVM) method, which can be used to anomaly detection of typical complex systems (Lyu et al., 2023; Fuertes et al., 2016). Gorinevsky et al. propose a real-time detection algorithm through estimating the hidden states and its effectiveness is verified with NASA public datasets (Gorinevsky et al., 2009). Biswas et al. propose an anomaly detection method based on unsupervised learning methods with simulation data (Biswas et al., 2016). And it can improve overall safety of the mission by data mining and analytics methods. Dong et al. present an anomaly detection method based on a LSTM model for typical complex systems, which has a high precision rate in different types of anomalies of the anomaly detection (Dong et al., 2019). Galal et al. present an anomaly detection method with a supervised Naïve Bayesian classifier model, which has achieved high detection accuracy on typical complex systems (Galal et al., 2019). And it overcomes the problem of the unavailability of labeled anomaly data in a supervised classification. Zhang et al. propose an improved adaptive Wiener process model based on sensing data with dynamic operating conditions to realize degradation estimation of electromechanical actuators (Zhang et al., 2022). Wang et al. present a deep learning domain adaptation method that introduces joint maximum mean discrepancy into convolutional neural network, which can improve the fault diagnosis accuracy in source and target domains (Wang et al., 2022b).

In addition, considering the time-series characteristic of the sensing data for complex systems, the prediction-based method can effectively improve the performance of PHM software platforms by combining offline data modeling and online prediction (Pan et al., 2015; Ma et al., 2015). Thus, it is widely used in the field of PHM software platform. For example, as a typical data-driven method, Gaussian Process Regression (GPR)-based prediction method can be used to achieve online detection of anomaly data samples and has a strong interpretation of the anomaly detection (Rasmussen, 2004). It is also a probabilistic prediction method with the ability of uncertainty expression and prediction intervals under arbitrary prediction confidence, which is more applicable in the field of complex system PHM software platforms.

On the other hand, due to the uncertainty of sensing data, the robustness of PHM software platform for complex systems needs to be evaluated to ensure its performance (Liu et al., 2022b). Generally, the good robustness mainly reflects the fact that the related methods can obtain a high detection accuracy with different signal-to-noise ratio noise input and is minimally affected by perturbations under many different operating conditions (Wang et al., 2022a). However, the sensing data are collected from complex systems may have distortion and noise data inevitably. Moreover, the variable operation environment aggravates the uncertainty of sensing data for complex systems (Biswas et al., 2016). These problems mentioned above lead to large uncertainties in the performance of PHM software platform for complex systems. Therefore, how to evaluate the robustness of PHM software platform for complex systems is challenging for researchers and engineers.

In current research, there are some lack of fair, objective and reasonable robustness evaluation methods for the PHM software platform of complex systems (Lyu et al., 2023; Liu et al., 2022a). In addition, as the number of robustness indicators increases, manual selection-based robustness evaluation methods will be highly subjective, time-consuming and inefficient. Therefore, it is critical to establish a fair, objective, and efficient robustness evaluation method for the PHM software platform of complex systems (Zeng et al., 2022). To solve these problems, in this paper, a novel adaptive robustness evaluation method for the PHM software platform of complex systems is proposed. In particular, "adaptive" is reflected in the fact that the algorithm can automatically optimize and extract more sensitive indicators as the number of indicators rises in practical engineering applications, and does not depend entirely on expert knowledge, which greatly enhances the efficiency of the PHM software platform. The algorithm is implemented as follows. Firstly, the basic robustness indicators are obtained for PHM software platform of complex systems to be evaluated, which include some common indicators and supplementary indicators. Meanwhile, on this basis, the mean and standard deviation of the basic robustness indicators are calculated. Secondly, the more sensitive indicators for PHM software platform of complex systems are selected by Genetic Algorithm (GA) from the basic indicators. In addition, the classification is conducted by the selected indicators through Random Forest Algorithm (RFA). Finally, the robustness evaluation results for PHM software platform of complex systems are determined by the classification accuracy.

The rest of the paper is organized as follows. Section 2 introduces that the basic theory including GPR, GA and RFA. Section 3 shows the anomaly detection framework and adaptive robustness evaluation method. Section 4 describes the case study including the experimental data and experimental results of robustness evaluation method. Section 5 draws the conclusions and future works of this research.

2. Theoretical foundations

2.1. Gaussian process regression theory

Prediction-based methods can effectively improve the performance of PHM software platform by combining offline data modeling and online prediction, which are widely used for complex systems (Liu et al., 2022b). As a statistical learning method in a Bayesian framework, GPR algorithm converts prior distributions into posterior models by training historical data to obtain probabilistically meaningful prediction results (Rasmussen, 2004). Due to the advantages such as simple training model, flexible nonparametric inference, and adaptive acquisition of hyperparameters, GPR has also become one of the most promising regression analysis methods with high robustness. Therefore, this paper takes the GPR-based PHM method as the specific object to prove the effectiveness of the proposed robustness evaluation method. In the GPR, it is assumed that the dependent variable y is composed of Gaussian distribution function $h(x)$ as well as noise ε, which is shown as follows.

$$y = h(x) + \varepsilon, \tag{1}$$

where ε follows a Gaussian distribution $\varepsilon \sim N\left(0, \sigma_n^2\right)$, and σ_n^2 is the variance of the noise distribution function. Besides, $h(x)$ is determined by expectation $m(x)$ as well as variance $k\left(x, x'\right)$, shown as follows.

$$h(x) \sim GP\left(m(x), k\left(x, x'\right)\right), \tag{2}$$

where $k\left(x, x'\right)$ indicates the covariance function also known as the kernel function. The solution of distribution function y can be obtained from $h(x)$ and ε, shown as follows.

$$y \sim GP\left(m(x), k\left(x, x'\right) + \sigma_n^2 \delta\left(x, x'\right)\right), \tag{3}$$

where $\delta\left(x, x'\right)$ is Kronecker delta function. If $x = x'$, then $\delta\left(x, x'\right) = 1$, otherwise $\delta\left(x, x'\right) = 0$.

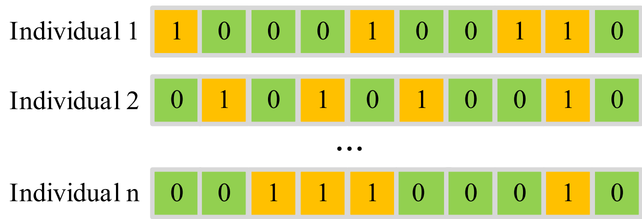


Fig. 1. Population individual code.

When the test data are input the model, the joint distribution of the training and test samples can be calculated as follows.

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{K}_*^T \\ \mathbf{K}_* & \mathbf{K}_{**} \end{bmatrix}\right), \tag{4}$$

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}, \tag{5}$$

where $\mathbf{K}$ is the symmetric positive definite covariance matrix with $n \times n$ order. $\mathbf{K}_*$ is the $1 \times n$ order covariance matrix between test point $\mathbf{x}_*$ and training set input, and $\mathbf{K}_{**}$ is the covariance of $\mathbf{x}_*$.

2.2. Basic principles of genetic algorithm

GA is an intelligent optimization algorithm based on natural selection and population genetic mechanism, which can simulate the natural evolutionary process of "Survival of the Fittest" in nature (Pei et al., 2022). It transforms the process of solving a problem into a population (i.e., a set of feasible solutions) selection optimization process. After reproduction, crossover, mutation and other basic operations, the population is selected to retain the better individuals. And through continuous iteration, it converges to the individual with the strongest ability to adapt to the environment (i.e., the optimal solution to the problem) (Fu et al., 2019). The components of GA are shown as follows.

(1) The encoding method

The binary encoding method is used in this paper and each chromosome is a string of zeros and ones together (Lambora et al., 2019). Where one means selected and zero means unselected. Besides, when the number of individuals is n, the population size consists of N strings of 0–1 corresponding to the value of the objective function. The composition of the population is shown in Fig. 1.

(2) Population initialization

The quality and number of feasible solutions can affect the iteration rate and optimization effect of the GA. The greater the number of feasible solutions, the greater the diversity of its population. And the less likely the resulting result will fall into a local optimum. But when the number is too excessive, it will significantly impact the convergence speed. Therefore, the number of populations should be chosen reasonably in the solution process according to the actual situation at the time of initialization.

(3) The fitness function

The natural selection process of GA is based on comparing the fitness value of each individual to determine whether the feasible solution is good or bad. The best projection direction based on the principle of differentiability of class-to-class distances is determined to ensure that the dispersion between classes is as large as possible in this paper. The average of the distances between any two samples in the two classes is used to represent the distance between the two classes. Besides, the distance-based method can be derived to the multiclass case. The specific analysis is as follows.

Let $\mathbf{x}_k^{(i)}, \mathbf{x}_l^{(j)}$ be the D-dimensional vectors in classes ϖ_i and ϖ_j, respectively, and $\delta\left(\mathbf{x}_k^{(i)}, \mathbf{x}_l^{(j)}\right)$ denotes the distance between these two vectors. Then the average distance between the vectors in each class is

$$J_d(x) = \frac{1}{2}\sum_{i=1}^{c} P_i \sum_{j=1}^{c} P_j \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \delta\left(x_k^{(i)}, x_l^{(j)}\right), \tag{6}$$

where c denotes the number of categories. n_i. and n_j are the number of samples in ϖ_i and ϖ_j categories, respectively. P_i, P_j are the prior probability of the corresponding category.

There are multiple distance measures between two variables in a multidimensional space that satisfy in the Euclidean distance case:

$$\delta\left(x_k^{(i)} + x_l^{(j)}\right) = \left(x_k^{(i)} - x_l^{(j)}\right)^T \left(x_k^{(i)} - x_l^{(j)}\right), \tag{7}$$

$$m_i = \frac{1}{n_i}\sum_{k=1}^{n_i} x_k^{(i)}, \tag{8}$$

$$m = \sum_{i=1}^{c} P_i m_i, \tag{9}$$

where m_i represents the mean amount of the sample set of the ith category, m denotes the total mean amount of the sample set of all categories. By bringing it into Eq. (6), we can get:

$$J_d(x) = \sum_{i=1}^{c} P_i \left[\frac{1}{n_i}\sum_{k=1}^{n_i}\left(x_k^{(i)} - m_i\right)^T \left(x_k^{(i)} - m_i\right) + (m_i - m)^T (m_i - m)\right]. \tag{10}$$

To ensure that the class-to-class dispersion is as large as possible and the intra-class dispersion is as small as possible, the distance criterion is defined as follows.

$$J_d(x) = \frac{trS_b}{trS_w}, \tag{11}$$

$$\widetilde{S}_b = \sum_{i=1}^{c} P_i (m_i - m)(m_i - m)^T, \tag{12}$$

$$\widetilde{S}_w = \sum_{i=1}^{c} P_i \frac{1}{n_i}\sum_{k=1}^{n_i}\left(x_k^{(i)} - m_i\right)\left(x_k^{(i)} - m_i\right)^T. \tag{13}$$

The definition of separability criterion based on intra-class distance is intuitive and easy to implement. And according to above equation, it is known that a larger value of $J_d(x)$ indicates better separability.

(4) Genetic operators

For iterative operations, the selection operator, the crossover operator and the variational operator are adopted in GA (Lambora et al., 2019). The most common selection operator is the roulette, in which the probability of an individual being selected is determined by the magnitude of its fitness value. And the larger the fitness value, the more likely the individual is to be selected.

The crossover operator is a two-by-two pairing of individuals in a population by first selecting the crossover position and then exchanging the chromosome segments at the selected position, which in turn generates two new ones. In this paper, a typical single-point crossover method is used, as shown in Fig. 2.

The mutation operator alters the genes in an individual to form a new one, thus expanding the scope of the solution. In nature,

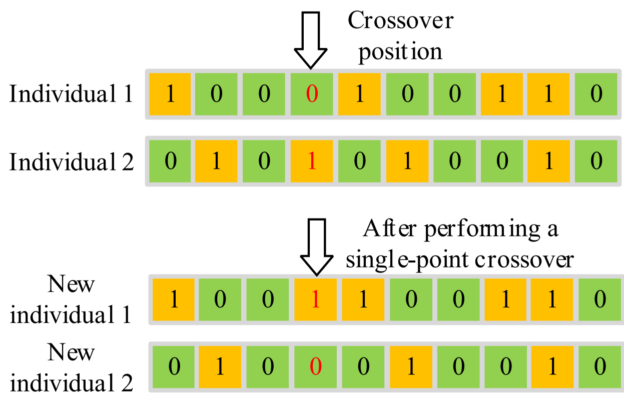


Fig. 2. Single point crossover strategy.

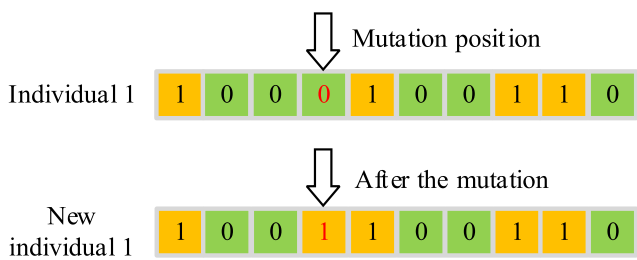


Fig. 3. Mutation operation.

genetic mutations have a low probability of occurring. When the mutation operation is done, it should be evaluated whether the individual should be mutated according to the specified mutation probability. And the mutation will be performed under certain rules, as shown in Fig. 3.

(5) Iterative operation

After the above operations, the population of the previous generation $M(t)$ becomes the population of the new generation $M(t+1)$. During the iteration of the GA, the chromosomes in the population converge to the optimal solution in the number of variables selected. And after reaching a certain number of iterations, the algorithm halts and outputs the chromosome with the highest fitness value in the population.

2.3. Random forest algorithm analysis

Random Forest Algorithm (RFA) is a parallel integrated learning method using decision trees as the base learner (Pavlov, 2000). In this case, the Decision Tree (DT) uses a tree structure and hierarchical inference to achieve the ultimate classification. The DT mainly consists of root nodes (i.e., containing the full set of samples), internal nodes (i.e., corresponding to feature attributes) and leaf nodes (i.e., decision results). Besides, there is no correlation between the DTs in the RFA, each DT is trained and operated independently simultaneously. When a new sample is input, the DTs used in the RFA are utilized to judge and classify the new samples, and the result is decided by each DT vote eventually.

3. Framework of the proposed methods

The robustness evaluation of PHM software platform for complex system in this paper can be divided into three main parts: PHM implementation, indicators calculation and adaptive robustness evaluation. The flowchart of this framework is shown in Fig. 4, and the details are described in the following subsections.

3.1. The principle of PHM implementation

As mentioned above, the original data are set by setting different noise parameters to obtain test multiple samples in Fig. 4, and then input to RFA. The second group noise is to obtain the composite metrics such as mean and variance of the GPR model. With two different sets of noise parameter settings, the model's resistance to disturbances can be tested and lay a foundation for robustness evaluation. The PHM software platform of complex systems mainly includes condition monitoring, anomaly detection, fault diagnosis, remaining useful life prediction and etc. On the other hand, anomaly detection is an important part of the reliable operation of complex system PHM software platforms, which enables the identification of abnormal behavior and timely takes actions to prevent system failures or accidents. Therefore, this paper takes the anomaly detection as an example to prove the effectiveness of the proposed robustness evaluation method. By examining the robustness of anomaly detection models, this study can provide valuable insights into the performance and limitations of PHM software platforms for complex systems. In turn, it can provide the development of more effective and reliable system health management strategies and support the safe and efficient operation of complex systems in various domains, such as aerospace, manufacturing, and transportation. Generally, the anomalies are categorized into four types: deviation anomaly, drift anomaly, burst anomaly, and static anomaly. Their mathematical descriptions are shown as follows (Wang and Wang, 2020; Wang et al., 2019, 2020; Zhong et al., 2022).

(a) Deviation anomaly and burst anomaly

For deviation anomalies and burst anomalies, which mean that the original data get an unexpected additional value k in a specific time period. The differences between them are that the additional value of the burst anomaly is generated by the random number and the duration of the additional values is quite short.

$$y_1(t)^* = \begin{cases} y_1(t) + k, \textit{ anomaly value} \\ y_1(t), \qquad \textit{others} \end{cases} \tag{14}$$

(b) Drift anomaly

Drift anomaly is the deviation of the original data from the normal value over time after the anomaly occurs.

$$y_2(t)^* = \begin{cases} y_2(t) + kt, \textit{ anomaly value} \\ y_2(t), \qquad \textit{others} \end{cases} \tag{15}$$

(c) Static anomaly

Static anomaly means that the original data become a static constant at the exception time point.

$$y_3(t)^* = \begin{cases} k, \textit{ anomaly value} \\ y_3(t), \qquad \textit{others} \end{cases} \tag{16}$$

where k is constant representing the degree of offset of the outlier. And $y_1(t)$, $y_2(t)$ and $y_3(t)$ are original data, $y_1(t)^*$, $y_2(t)^*$ and $y_3(t)^*$ are anomaly data.

Besides, the anomaly data are obtained from the empirical formula, and the anomaly proportion is 10%, as shown in Fig. 5, where (a), (c), (e), (g) represent deviation anomaly, drift anomaly, static anomaly and burst anomaly data respectively, and the remaining ones represent their corresponding labels. And the red, blue, and black curves indicate anomaly data, normal data, and anomaly labels respectively.

The GPR algorithm is used for anomaly detection with the preprocessed raw data. Then the mean function, covariance function and hyperparameters of GPR model are initialized. On this basis, the former 400 data points are adopted for training and the last 600 data points are tested for prediction. Besides, the 3σ

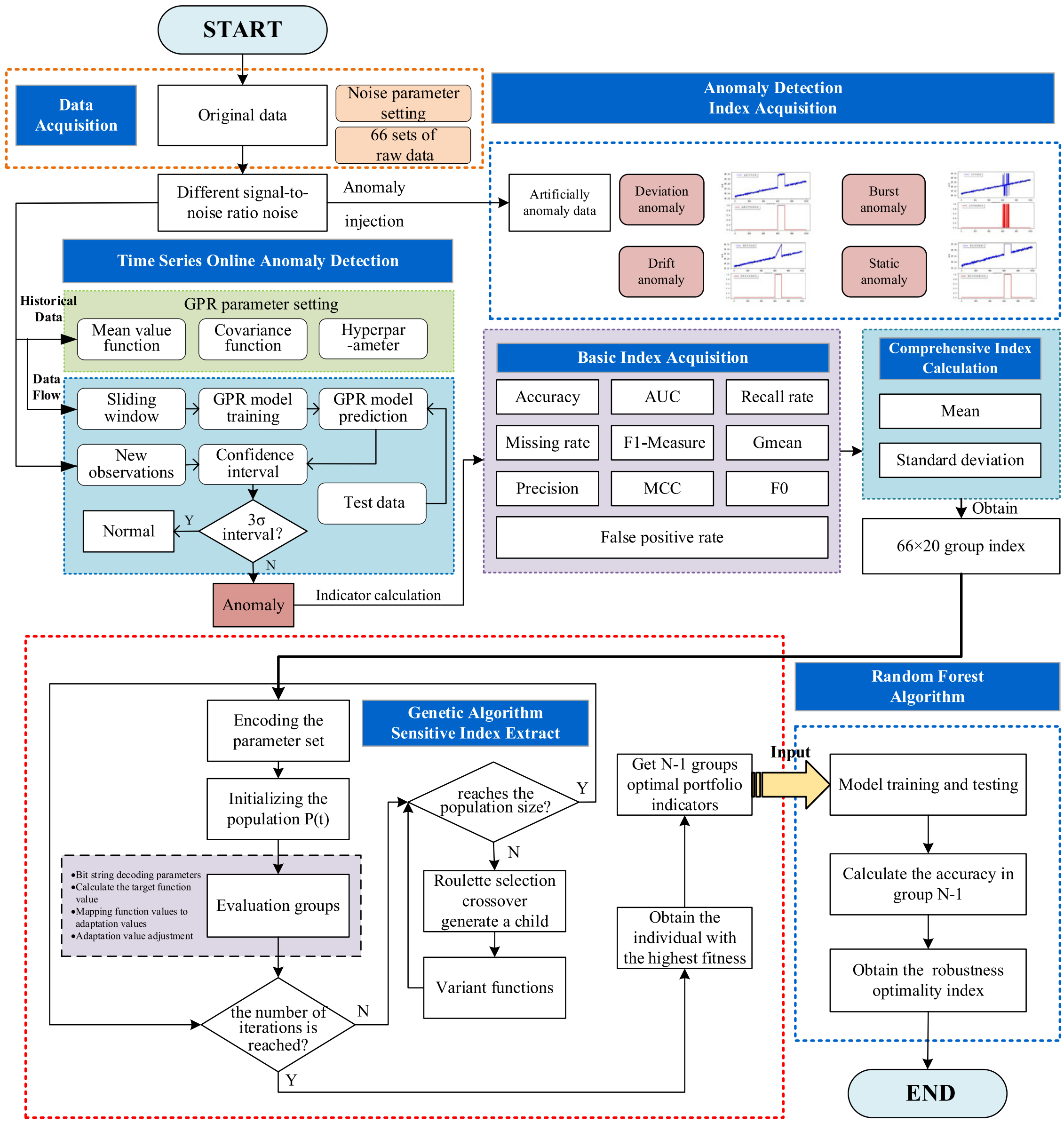


Fig. 4. The flowchart of the proposed method (anomaly detection).

confidence interval $(\Pr(\mu - 3\sigma \le X \le \mu - 3\sigma) \approx 0.9973)$ is used as the uncertainty range of the predicted data. And μ, σ represent the mean and variance, respectively. If the input data exceeds this range, it is determined to be an anomaly.

3.2. Indicators calculation principle

In this paper, the process of the elementary indicator calculation adopts the metrics commonly used in the anomaly detection methods such as accuracy rate, AUC, F1-Measure, recall rate and etc. In addition, Matthews Correlation Coefficient (MCC) (Boughorbel et al., 2017), F0 (Ji et al., 2021), and G-mean (Wang and Yao, 2012) are used as supplementary indicators, which can be used to reflect the performance of the model. On this basis, their means and standard deviation are calculated separately, which reflect the quantity and dispersion of the concentration trend of each indicator, respectively. So, the composite indicators are obtained on the foundation of the basic indicators.

Generally, the performance of the anomaly detection methods is analyzed using a confusion matrix, as shown in Table 1. TP indicates that the sample of the actually positive class is predicted as a positive class. FN represents that the sample of the actually positive class is predicted as a negative class. FP indicates that the sample of the actually negative class is predicted as a positive class. TN represents that the sample of the actually negative class is predicted as a negative class.

Table 1
Confusion matrix.

Real data	Predicted Data	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

Based on the confusion matrix, the mathematical descriptions for the basic indicators used in this paper, are shown in Table 2.

M is the number of the positive class, and N is the number of the negative class. $\sum_{i \in positiveClass} rank_i$ represents the summation after sorting of all positive examples. And the f_i denotes the overlapping degree of ith anomaly subsequence between

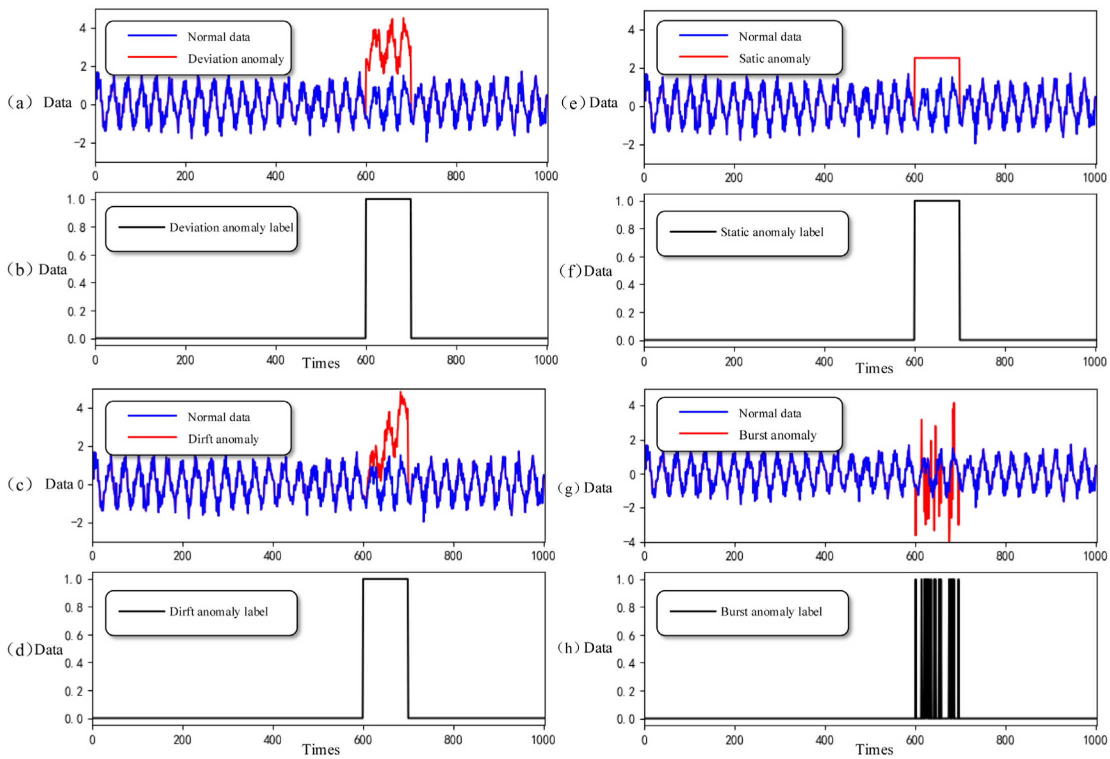


Fig. 5. Example analysis of anomaly data.

Table 2
Formula for indicator calculation.

Symbols meaning	Formula	Symbols meaning	Formula
Accuracy	$acc = \frac{TP+TN}{TP+TN+FP+FN}$	AUC	$AUC = \frac{\sum_{i \in positiveClass} rank_i - \frac{M(1+M)}{2}}{M \times N}$
False Positive Rate	$fp = \frac{FP}{TN+FP}$	MCC	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
Missing Rate	$miss = \frac{FN}{TP+FN}$	F0	$F0 = \frac{1}{N}\sum_{i=1}^{N} f_i,\ f_i = \frac{length(Pre_i - Ref_i)}{length(Ref_i)}$
Recall Rate	$recall = \frac{TP}{TP+FN}$	G-mean	$.G = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}}.$
Precision	$pre = \frac{TP}{TP+FP}$	Mean	$mean = \frac{x_1+x_2+\cdots x_n}{n}$
F1-Measure	$F1 = \frac{2TP}{2TP+FN+FP}$	Standard Deviation	$std = \sqrt{\frac{1}{n}\left[(x_1-\bar{x})^2 + \cdots (x_n-\bar{x})^2\right]}$

prediction (Pre_i) and observation (Ref_i); $x_1, x_2, \ldots, x_n$ mean the values of indicators at the points of $1, \ldots, n$. $\bar{x}$ indicates the mean of $x_1, x_2, \ldots, x_n$ and n is the number of points. Besides, the lower values of the missing rate and fault positive rate indicate better algorithm performance, and higher values of the remaining metrics indicate better robustness performance.

3.3. The principle of adaptive robustness evaluation

All the indicators (i.e., elementary and composite indicators) are obtained from Section 3.2. And the adaptive robustness evaluation method includes two parts: optimal indicators selecting (i.e., GA) and classification accuracy calculating (i.e., RFA). Among them, the optimal indicators refer to the indicators obtained by GA. Firstly, the parameter set encoder, the population initialization and the groups evaluation need to be finished in GA. Then determine whether the model has reached the number of iterations and the population size. If both not, the operations including the roulette selection crossover and variant functions setting should be adopted. On this basis, it can obtain the individual with the highest fitness (i.e., optimal indicators). Next, the above obtained indicators are divided into training and test sets and input into the random forest model. Furthermore, good robustness mainly reflects the fact that the related methods can obtain a high detection accuracy with different signal-to-noise ratio noise input and is minimally affected by perturbations under many different operating conditions. And "classification accuracy" is used to reflect the robustness performance by setting parameters with different signal-to-noise ratios. The better robustness indicates that the samples are less affected by interference at different signal-to-noise ratios, and the samples are more difficult to be distinguished from each other (i.e., the separation is more difficult). According to the above analysis, by calculating the classification accuracy, the result of robustness assessment can be obtained.

4. Case study

4.1. Experimental data description

As one of the core components of spacecraft, the spacecraft power system is used to generate, store, transform, regulate and distribute electrical energy, which is a typical complex system.

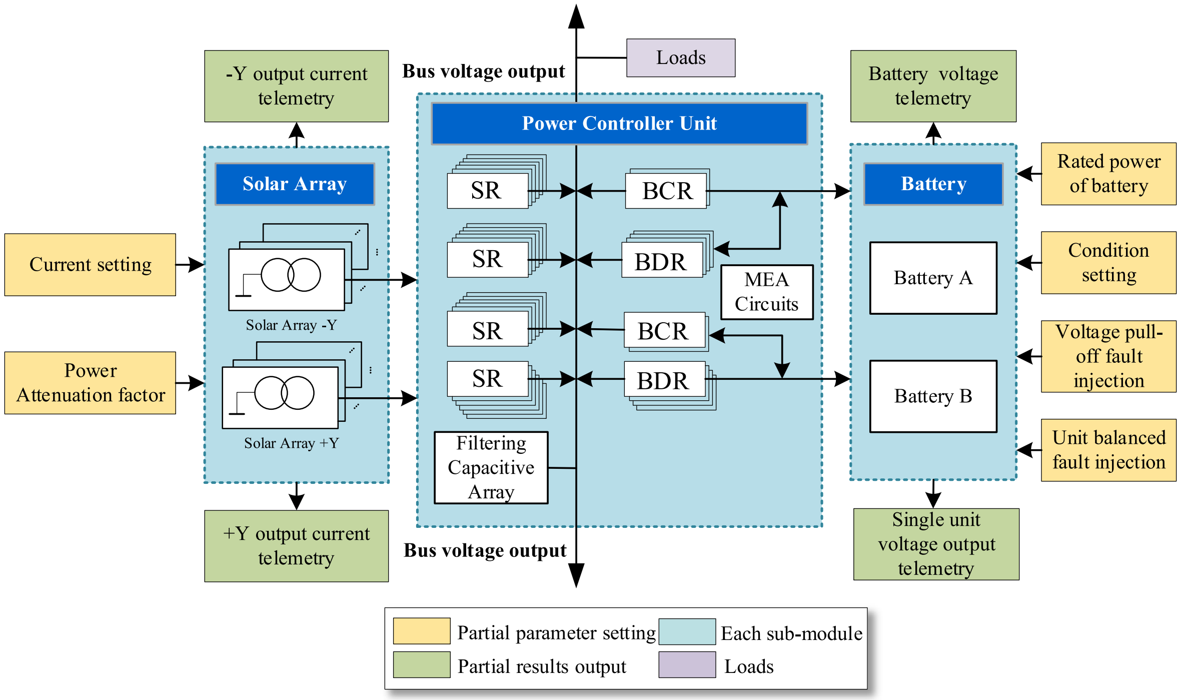


Fig. 6. The structure of spacecraft power system.

Therefore, in this paper, to prove the effectiveness of the proposed method, a reliable simulation model of spacecraft power system is established by analyzing the working principle. The data of this experiment are obtained from the simulation model, as shown in Fig. 6. The model mainly includes three parts: solar array (SA), battery and power control unit (PCU).

When the spacecraft enters the illumination area, the solar panels generate electrical energy through the photovoltaic effect. During the shadow area, the voltage of the SA decreases gradually as the illumination decreases. At this time, the spacecraft is powered entirely by the battery set. The PCU consists of the shunt regulator (SR), battery charge regulator (BCR) and battery discharge regulator (BDR). PCU is mainly adopted to realize the functions of regulating the output power of the SA, stabilizing the bus voltage as well as controlling the battery bank for safe charging and discharging, and etc. In addition, the main error amplifier (MEA) circuit module provides the bus error amplification signal to control the operating state of the SR, BCR, and BDR modules.

4.2. PHM implementation

As mentioned above, anomaly detection is used as an example to prove the effectiveness of the proposed robustness evaluation method. Considering the importance of bus voltage stability in the normal operation of spacecraft power systems, the bus voltage output of the simulation model is adopted for anomaly detection in this paper. And it is characterized by single-dimensional data. The results of deviation anomaly detection, drift anomaly detection, burst anomaly, and static anomaly are analyzed as shown from Figs. 7 to 10, where the pink, black, green, and blue curves represent the offline training data, the anomaly data, the location of the actual anomaly implantation, and the prediction results output by the GPR model, respectively. Besides, the light pink area with yellow dotted line represents the confidence interval and the red ones represent the actual detected anomalies. It can be noted from Fig. 7 that all of the injected deviation anomalies are detected, but a few false alarms still exist. In Fig. 8, due to the threshold setting characteristics, the drift anomalies have few data which is still judged to be normal. Besides, in the process of burst anomaly detection and static anomaly detection, there are some false alarms as the noise exceeds the threshold.

Table 3
The parameters of the GA.

Number of the population	100
Number of the iteration	100
Mutation probability	0.002
Crossover probability	0.6

4.3. Adaptive robustness evaluation

The means and standard deviations of all elementary indicators are available after anomaly detection in the previous subsection, which are also utilized as input to the adaptive robustness evaluation method. And partial indicators of deviation anomaly detection calculation results are shown in Fig. 11.

In Fig. 11, the mean and standard deviation of accuracy, ROC, false positive rate, missing rate, F1-Measure, recall rate, precision, MCC, G-mean, and F0 are shown in order. The "Number" represents the experiment number for each indicator. In addition, the "Value" represents the values of the different indicators.

Based on the calculated indicators, GA is used to adaptively select the optimal indicators, which represent more sensitive indicators of system robustness performance. The combination of the obtained indicators is separately treated as the input of the Random Forest Algorithm (RFA) to calculate the classification accuracy as the basis of robustness assessment. The parameters of the GA in this experiment are set as shown in Table 3. And the selected sensitive indicators through GA are shown in Table 4. And the number of the extracted indicators of deviation anomaly, drift anomaly, burst anomaly, and static anomaly is 13, 7, 17, and 10, respectively.

In the RFA, all selected sensitive indicators are divided into five categories according to the different types of the input noise, and the test and training sets are divided randomly in a 1:1 ratio.

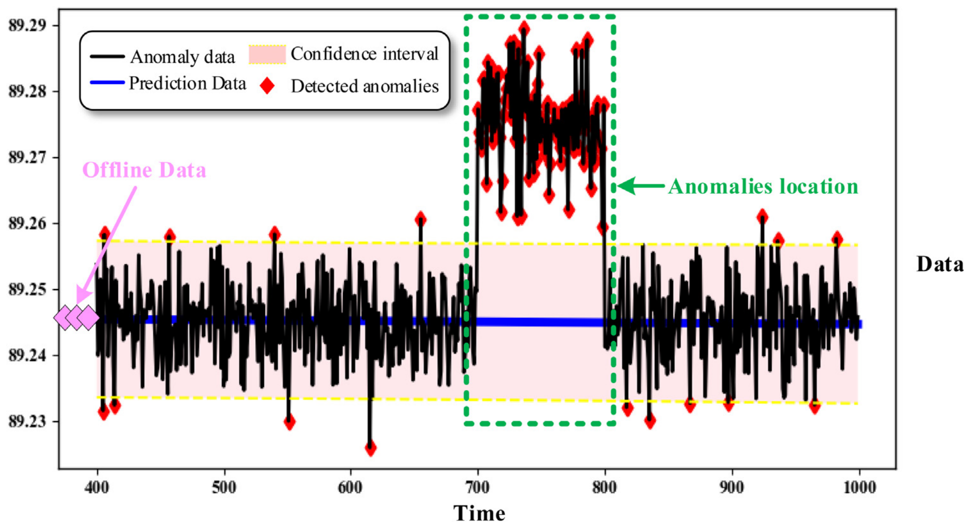


Fig. 7. Deviation anomaly detection.

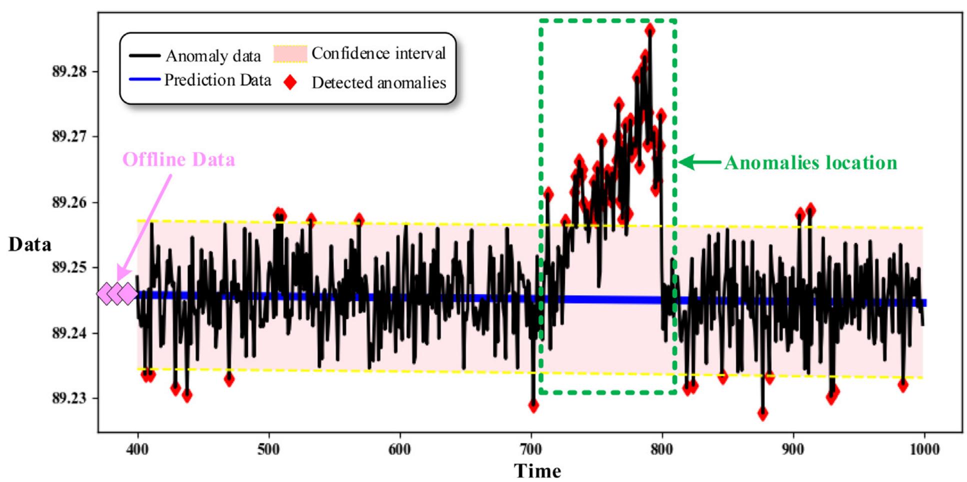


Fig. 8. Drift anomaly detection.

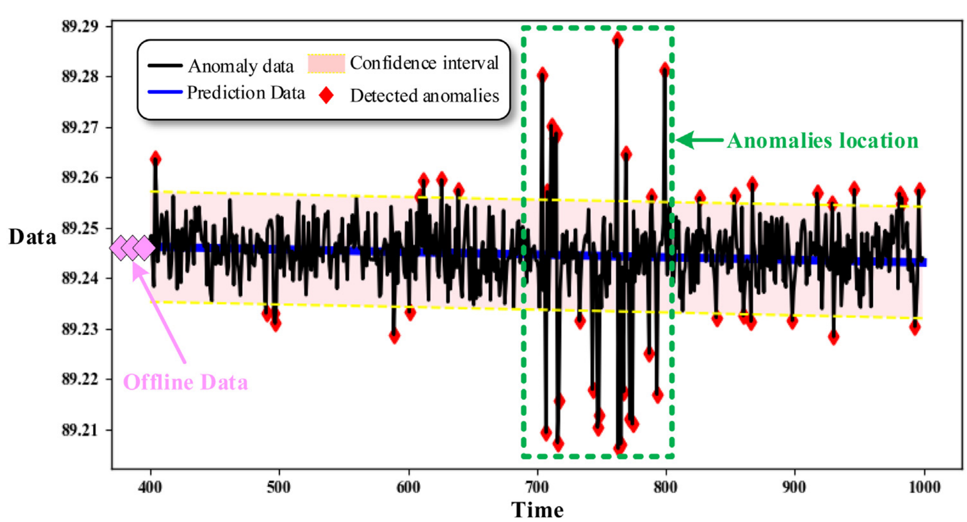


Fig. 9. Burst anomaly detection.

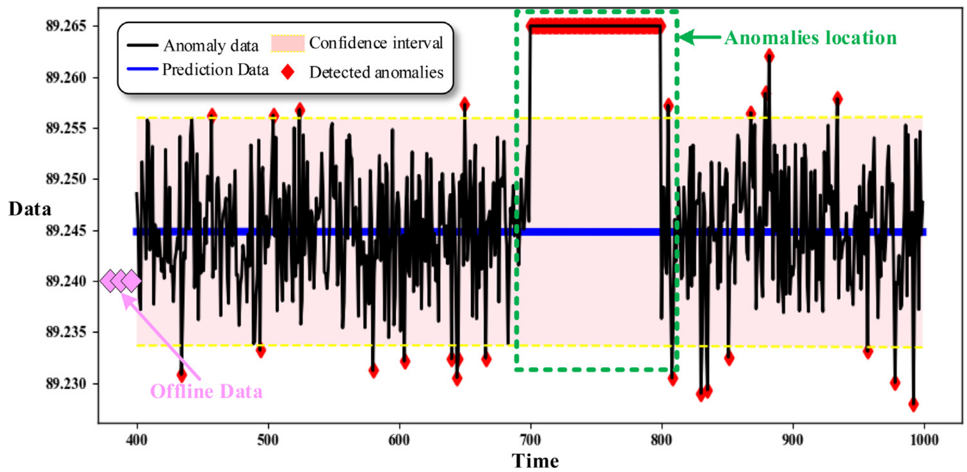


Fig. 10. Static anomaly detection.

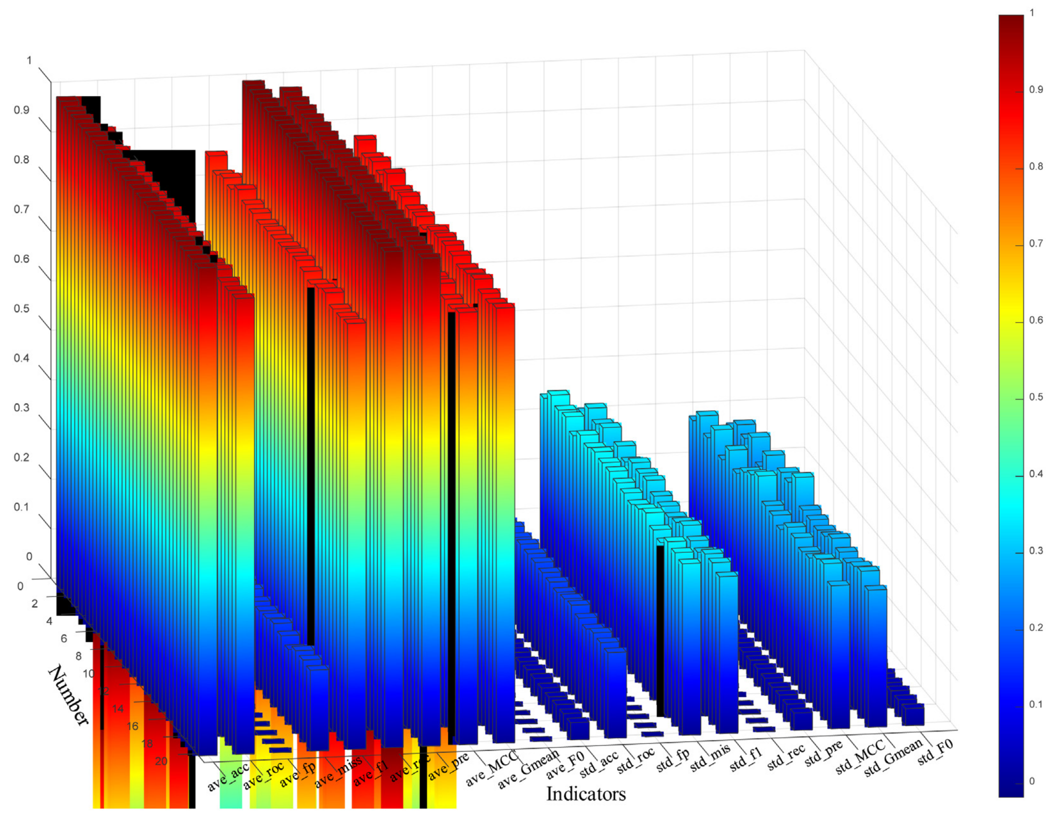


Fig. 11. Visualization of partial indicators of deviation anomaly results.

4.4. comparison and analysis

Due to the advantages of low computational complexity and high interpretability, many machine learning methods are currently being applied by many researchers for performance verification of robustness evaluation, such as Support Vector Machine (SVM) and Decision Tree (DT). Among them, SVM has good robustness for learning from small sample datasets and effectively avoids overfitting. On the other hand, DT has some advantages in terms of feature selection. In general, SVM and DT have high accuracy and reliability, which are widely used in robustness evaluation and showing good prospects for practical applications. In this paper, SVM and DT are used as comparison algorithms for validation of robustness evaluation. The robustness results using different algorithms for the four anomalies are shown in Table 5.

In Table 5, the robustness evaluation result is represented the "1-classification accuracy" in this experiment, which indicates that the closer the value is to one, the better the robustness

Table 4
The selected sensitive indicators by GA.

Number	Anomaly type	Selected indicators
1	Deviation anomaly	ave_acc, ave_roc, ave_rec, ave_pre, ave_Gmean, ave_F0, std_acc, std_roc, std_fp, std_miss, std_f1, std_MCC, std_F0;
2	Drift anomaly	ave_rec, ave_pre, ave_Gmean, std_acc, std_roc, std_fp, std_MCC;
3	Burst anomaly	ave_acc, ave_roc, ave_fp, ave_miss, ave_f1, ave_rec ave_pre, ave_MCC, ave_Gmean, ave_F0, std_acc, std_roc, std_fp, std_miss, std_f1, std_MCC, std_Gmean;
4	Static anomaly	ave_roc, ave_rec, ave_pre, ave_Gmean, std_acc, std_roc, std_fp, std_miss, std_f1, std_MCC.

performance. And the "optimization percentage" denotes the degree of optimization of the proposed method over the other two methods in terms of "1-accuracy". For deviation anomaly, the proposed method first selects 13 more sensitive indicators by GA and input them into the RFA for classification, and the "1-classification accuracy" is 0.96970 in this experiment. Meanwhile, SVM and DT use all 20 sets composite indicators for classification, and the robustness results obtained were 0.90909 and 0.87273, respectively. Similarly, for the drift anomaly, the proposed algorithm has a robustness result of 0.93939, and the robustness results for SVM and DT are 0.78788 and 0.90909, respectively. The optimization percentage can be calculated as follows.

$$OP = \frac{PM - OA}{OA} \times 100\% \tag{17}$$

$$ave\_OP = \frac{OP_1 + OP_2 + OP_3 + OP_4}{4} \times 100\% \tag{18}$$

where OP, PM, OA indicate the "optimization percentage", "the results of the proposed method" and "the results of the other algorithms", respectively. And the ave_OP indicates the average value of the OP. OP_1, OP_2, OP_3, OP_4 represent the OP value of different anomalies. By calculating its optimization percentage through the formula, it is possible to compare the advantages and disadvantages between algorithms under different anomalies. It can draw a conclusion from Table 5 that the algorithm used in this paper has better robustness evaluation performance in deviation anomaly, drift anomaly, and static anomaly compared with SVM and DT algorithms, and the average optimization percentages of the evaluation results are 12.24% and 1.11%, respectively. Furthermore, different indicators can capture different aspects of the model's performance and behavior under different conditions, and each indicator may have its own strengths and limitations. Therefore, using a combination of indicators can provide a more comprehensive evaluation of the model's robustness. In this paper, the more sensitive indicators extracted by GA have better robustness results in classification.

However, due to the higher discrimination among different indicators of burst anomaly, the proposed method in this paper has lower robustness evaluation results on burst anomaly detection. It may be caused by that the GPR-based anomaly detection utilized in this paper is not suitable for burst anomaly. In summary, through the case study of GPR-based anomaly detection of spacecraft power systems, it can be concluded that the proposed method can be used for robustness evaluation for the PHM software platform of complex systems and has better adaptability.

5. Conclusions

This work focuses the robustness evaluation for the PHM software platform of complex systems with uncertainty of sensing data. To solve the problem of selecting multiple indicators as well as the strong subjectivity relying on expert experience in the robustness evaluation of complex system PHM software platform, a novel adaptive robustness evaluation method is proposed in this paper. Through the GA-based indicator selection and RFA-based classification, the proposed methods can implement the adaptive robustness evaluation for the PHM software platform of complex systems.

To prove the effectiveness of the proposed method, a case study of GPR-based anomaly detection is implemented in which a simulation model of the spacecraft power system is established. Besides, experiments with simulation data based on the proposed method and two comparison methods (i.e., SVM and DT) are conducted. Experimental results show that the proposed method has better performance in the robustness evaluation of deviation anomaly, drift anomaly and static anomaly for the spacecraft power system. Compared with SVM and DT methods, the average optimization percentages of the evaluation results are 11.39% and 3.33%, respectively. Thus, the proposed method has better adaptability and is suitable for the robustness evaluation of complex system PHM software platform. However, the performance of robustness evaluation on the burst anomaly detection of spacecraft power system is relatively poor.

In future work, more attention to the robustness evaluation of burst anomaly detection in PHM software platform of complex systems will be considered. In addition, facing a large of data in the future, deep learning algorithms are also gradually being

Table 5
Multiple algorithms for the robustness evaluation results.

Algorithms	Number of indicators	Anomaly type	Results	Optimization percentage
The proposed method	13	Deviation anomaly	**0.96970**	/
SVM	20		0.90909	6.67%
DT			0.87273	11.11%
The proposed method	7	Drift anomaly	**0.93939**	/
SVM	20		0.78788	19.23%
DT			0.90909	3.33%
The proposed method	17	Burst anomaly	0.81818	/
SVM	20		0.78789	3.84%
DT			**0.90909**	−10%
The proposed method	10	Static anomaly	**0.93939**	/
SVM	20		0.78788	19.23%
DT			**0.93939**	0

studied by many researchers. And long short-term memory has better capability in processing time series data compared to recurrent neural networks. And it will be the research directions for robustness evaluation algorithms in the future.

CRediT authorship contribution statement

He Liu: Carried out the experiments, Wrote the manuscript, Discussed the results, Contributed to the final manuscript. **Cheng Wei:** Conceived of the presented idea, Developed the framework of this study, Discussed the results, Contributed to the final manuscript. **Bo Sun:** Conceived of the presented idea, Developed the framework of this study, Discussed the results, Contributed to the final manuscript. **Yinxue Zeng:** Reviewed and edited the manuscript, Discussed the results, Contributed to the final manuscript.

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

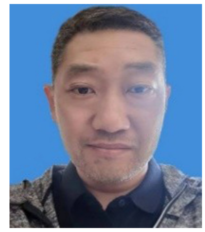
The data that has been used is confidential.

References

Biswas, G., Khorasgani, H., Stanje, G., Dubey, A., Deb, S., Ghoshal, S., 2016. An approach to mode and anomaly detection with spacecraft telemetry data. Int. J. Prognost. Health Manag. 7 (4), 1–18.
Boughorbel, S., Jarray, F., El-Anbari, M., 2017. Optimal classifier for imbalanced data using matthews correlation coefficient metric. PLOS ONE 12 (6), e0177678.
Dong, J., Ma, Y., Liu, D., 2019. Deep learning based multiple sensors monitoring and abnormal discovery for satellite power system. In: International Conference on Sensing, Diagnostics, Prognostics, and Control. SDPC, IEEE, pp. 638–643.
Fernandez, M.M., Yue, Y., Weber, R., 2017. Telemetry anomaly detection system using machine learning to streamline mission operations. In: International Conference on Space Mission Challenges for Information Technology. pp. 70–75.
Fu, R., Al-Absi, M.A., Abdulhakim Al-Absi, A., Lee, H.J., 2019. A conservation genetic algorithm for optimization of the E-commerce logistics distribution path. In: International Conference on Advanced Communication Technology. ICACT, IEEE, pp. 558–562.
Fuertes, S., Picart, G., Tourneret, J.Y., Chaari, L., Ferrai, A., Richard, C., 2016. Improving spacecraft health monitoring with automatic anomaly detection techniques. In: International Conference on Space Operations. pp. 1–16.
Galal, M.A., Hussein, W.M., El-din abdel Kawy, E., Sayed, M.M.A., 2019. Satellite battery fault detection using naïve Bayesian classifier. In: Aerospace Conference. IEEE, pp. 1–11.
Gorinevsky, D., Boyd, S., Poll, S., 2009. Estimation of faults in DC electrical power system. In: Proceedings of the American Control Conference. IEEE, pp. 4334–4339.
Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T., 2018. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: International Conference on Knowledge Discovery and Data Mining. pp. 387–395.
Ibrahim, S.K., Ahmed, A., Zeidan, M.A.E., Ziedan, I.E., 2019. Machine learning methods for spacecraft telemetry mining. IEEE T. Aero. Elec. Sys. 55 (4), 1816–1827.
Ji, Z., Gong, J., Feng, J., 2021. A novel deep learning approach for anomaly detection of time series data. Sci. Program. 2021, 1–11.
Ji, X., Li, Y., Liu, G., Guo, J., Xiang, S., Yang, X., 2019. A brief review of ground and flight failures of Chinese spacecraft. Prog. Aerosp. Sci. 107, 19–29.
Jiang, H., Zhang, K., Wang, J., Wang, X., Huang, P., 2021. Anomaly detection and identification in satellite telemetry data based on pseudo-period. Appl. Surf. Sci. 10 (1), 1–20.
Lambora, A., Gupta, K., Chopra, K., 2019. Genetic algorithm-a literature review. In: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing. IEEE, pp. 380–384.
Li, C., Ru, L., 2019. Prognostics and health management techniques for integrated avionics systems. In: IEEE Prognostics and System Health Management Conference. pp. 1–5.
Liu, B., Zhao, Y., Cao, Y., Kang, Y., Xu, Z., 2022a. A robustness benchmark for prognostics and health management. In: Chinese Control Conference. IEEE, pp. 7338–7342.
Liu, B., Zhao, Y., Cao, Y., et al., 2022b. A robustness benchmark for prognostics and health management. In: Chinese Control Conference. CCC, IEEE, pp. 7338–7342.
Lyu, G., Zhang, H., Miao, Q., 2023. RUL prediction of lithium-ion battery in early-cycle stage based on similar sample fusion under Lebesgue sampling framework. IEEE Trans. Instrum. Meas. 72, 1–11.
Lyu, G., Zhang, H., Zhang, Y., Miao, Q., 2022. An interpretable remaining useful life prediction scheme of lithium-ion battery considering capacity regeneration. Microelectron. Reliab. 138, 1–6.
Ma, J., Le, S., Wang, H., Zhang, Y., U., Aickelin., 2015. Supervised anomaly detection in uncertain pseudoperiodic data streams. ACM T. Internet. Techn. 16 (1), 1–20.
Pan, D., Liu, D., Zhou, J., Zhang, G., 2015. Anomaly detection for satellite power subsystem with associated rules based on kernel principal component analysis. Microelectron. Reliab. 55 (9–10), 2082–2086.
Pavlov, Y., 2000. Random Forests. De Gruyter, Berlin, Boston.
Pei, M., Goodman, E.D., Punch, W.F., Ding, Y., 2022. Genetic algorithms for classification and feature extraction. In: Classification Society Conference. pp. 355–7516.
Rasmussen, C.E., 2004. Gaussian processes for machine learning. In: Machine Learning Summer School Conference. pp. 63–71.
Soualhi, A., Lamraoui, M., Elyousfi, B., Razik, H., 2022. PHM SURVEY: Implementation of prognostic methods for monitoring industrial systems. Energies 15 (19), 6909.
Vogl, G.W., Weiss, B.A., Helu, M., 2019. A review of diagnostic and prognostic capabilities and best practices for manufacturing. J. Intell. Manuf. 30 (1), 79–95.
Wang, Y., Bai, G., Zhang, Y.A., et al., 2022a. Cascading failure analysis and robustness assessment of the operational system and electric power system based on dependent network. In: 2022 Global Reliability and Prognostics and Health Management. PHM-Yantai, IEEE, pp. 1–8.
Wang, B., Liu, D., Peng, Y., Peng, X., 2020. Multivariate regression-based fault detection and recovery of UAV flight data. IEEE Trans. Instrum. Meas. 69 (6), 3527–3537.
Wang, B., Liu, D., Peng, X., Wang, Z., 2019. Data-driven anomaly detection of UAV based on multimodal regression model. In: International Instrumentation and Measurement Technology Conference. I2MTC, IEEE, pp. 1–6.
Wang, X., Wang, C., 2020. Time series data cleaning with regular and irregular time intervals. pp. 1–17, arXiv preprint arXiv:2004,08284.
Wang, S., Yao, X., 2012. Multiclass imbalance problems: Analysis and potential solutions. Trans. Syst. Man. Cybern. Syst. 42 (4), 1119–1130.
Wang, J., Zhang, Y., Luo, C., Miao, Q., 2022b. Deep learning domain adaptation for electro-mechanical actuator fault diagnosis under variable driving waveforms. IEEE Sens. J. 22 (11), 10783–10793.
Wong, W.E., Debroy, V., Surampudi, A., Kim, H., Siok, M.F., 2010. Recent catastrophic accidents: Investigating how software was responsible. In: International Conference on Secure Software Integration and Reliability Improvement. IEEE, pp. 14–22.
Wong, W.E., Li, X., Laplante, P.A., 2017. Be more familiar with our enemies and pave the way forward: A review of the roles bugs played in software failures. J. Syst. Softw. 133, 68–94.
Yairi, T., Oda, T., Nakajima, Y., Miura, N., Takata, N., 2014. Evaluation testing of learning-based telemetry monitoring and anomaly detection system in SDS-4 operation. In: International Symposium on Artificial Intelligence, Robotics and Automation in Space. I-SAIRAS, pp. 1–8.
Yang, L., Ma, Y., Zeng, F., Peng, X., Liu, D., 2021. Improved deep learning based telemetry data anomaly detection to enhance spacecraft operation reliability. Microelectron. Reliab. 126, 1–6.
Zeng, Z., Guang, J., Chi, X., Chen, S, Zhang, L., 2022. Spacecraft telemetry anomaly detection based on parametric causality and double-criteria drift streaming peaks over threshold. Appl. Sci.-Basel 12 (4), 1–24.
Zhang, Y., Liu, D., Miao, Q., Peng, Y., 2022. Sensing data-based degradation estimation of electromechanical actuator under dynamic operating conditions. IEEE Sens. J. 22 (22), 21837–21845.
Zhang, P., Wu, M., Xiong. Present Situation, F., 2020. Challenge and future of health management about complex equipment. J. Phys. Conf. Ser. 1673, 012004.
Zhong, J., Zhang, Y., Wang, J., Luo, C., Miao, Q., 2022. Unmanned aerial vehicle flight data anomaly detection and recovery prediction based on spatio-temporal correlation. IEEE Trans. Reliab. 1 (1), 457–468.

He Liu received the B.S. degree from Harbin Institute of Technology (HIT), Harbin, China, in 2005, and the M.E. degree from Harbin Institute of Technology (HIT), Harbin, China, in 2007. He is currently a research fellow of Beijing Institute of Space System Engineering, CAST, Beijing, China. His current research interests include spacecraft integrated test and data processing, unattended test monitoring, spacecraft flight data simulation and analysis, machine learning, fault diagnosis.

Cheng Wei received the B.S. degree from Harbin Institute of Technology, Harbin, China, in 2005, and the Ph.D. degree from Harbin Institute of Technology (HIT), Harbin, China, in 2010. He is a professor with the School of Aerospace, Harbin Institute of Technology (HIT), Harbin, China. His current research interests include multibody system dynamics simulation, data-driven modeling and spacecraft digital twin technology.



Bo Sun received the B.S. degree from Harbin Institute of Technology, China, and the Ph.D. degree from Beihang University, Beijing, China. He is currently a professor with the Beijing Institute of Space Systems Engineering, Beijing, China. His current research interests include prognostics and health management, intelligent equipment testing, and model-driven digital twins.

Yinxue Zeng received the B.E. degree from the School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, China, in 2017. She is currently studying in College of Electrical Engineering, Sichuan University, Chengdu, China. Her current research interests include prognostics and health management of complex electromechanical systems, as well as anomaly detection, and credible evaluation modeling.