IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 5, 2019, accepted December 4, 2019, date of publication December 23, 2019,
date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961742

Steady-State Process Fault Detection for Liquid Rocket Engines Based on Convolutional Auto-Encoder and One-Class Support Vector Machine

XIAOBIN ZHU, YUQIANG CHENG, JIANJUN WU, RUNSHENG HU, AND XING CUI
College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China

Corresponding authors: Yuqiang Cheng (cheng_yuqiang@163.com) and Jianjun Wu (jjwu@nudt.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 50276068 and Grant 51206181.

ABSTRACT Liquid rocket engines (LREs) are the main propulsive devices of launch vehicles. Due to the complex structures and extreme working environments, LREs are also the components prone to failure. It is of great engineering significance to develop fault detection technologies which can detect fault symptoms in time and provide criteria for further fault diagnosis and control measures to avoid serious consequences during both the ground tests and flight missions. This paper presents a novel fault detection method based on convolutional auto-encoder (CAE) and one-class support vector machine (OCSVM) for the steady-state process of LREs. We train the CAEs by normal ground hot-fire test data of a certain type of large LRE for automatic feature extraction. Then the obtained features are used to train the OCSVMs to accomplish the fault detection task. The results demonstrate that the proposed method outperforms traditional redline system (RS), adaptive threshold algorithm (ATA), and back-propagation neural network (BPNN). We also study the effect of sample sizes and domain knowledge on the performance of the proposed method. The results suggest that appropriate measures that enrich the effective information content in the training data, such as increasing sample size and introducing domain knowledge, can further improve the performance of the proposed fault detection method.

INDEX TERMS Convolutional auto-encoder, one-class support vector machine, liquid rocket engine, steady-state process, fault detection.

I. INTRODUCTION

As an irreplaceable propulsive device, the liquid rocket engine (LRE) is one of the most critical components of the launch vehicle. The complex structures and extreme working environments make LREs the most vulnerable subassemblies and prone to failure [1]. Since the 1960s, with the development of major projects with milestone significance, such as manned moon landings and space shuttles, more attention has been paid to the reliability and security of LREs [2], [3]. Currently, health monitoring technology has become one of the important and effective means to ensure the safe and reliable operation of LREs. And the fault detection theories and methods, as the basis of health monitoring, have always been the research focus in this field [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Imran Sarwar Bajwa.

At present, fault detection algorithms for LREs can basically be classified into the threshold detection methods, whose key research contents are the generation of monitoring indicators and the determination of thresholds, among which the former mainly includes the methods based on mathematical model and signal processing, and the latter generally relies on expertise and engineering experience. Meanwhile, with the development of data mining and machine learning (ML), related fault detection algorithms are also emerging [5]–[9]. However, almost all the above methods need the selection of monitoring parameters and the extraction of artificial designed fault features from the original data to effectively characterize the engine faults, which is usually time-consuming, labor-intensive, less portable, and highly depend on domain knowledge [10].

Since proposed by Hinton and Salakhutdinov [11] and Hinton *et al.* [12] in 2006, deep learning (DL) has got much

attention and been widely studied in many fields, such as science, business, and government [13]. Many breakthroughs have been achieved with the help of DL in various tasks, e.g., image recognition [14], [15], speech recognition [16] and natural language processing [17], [18]. In recent years, the significant advantages of DL in automatic feature extraction of high-dimensional data have gradually attracted the attention of the community of system health monitoring (SHM). Though some attempts have been made [19]–[23], the research of DL in SHM is still preliminary. Most of the studies focus on the fault detection, diagnosis and performance degradation assessment of some critical components, e.g., bearings and gearboxes [24]. There is little applied research on DL for the health monitoring of complex systems, especially in the aerospace industry. Cortes and Rabelo [25] conducted a preliminary study on the architecture for monitoring and anomaly detection for space systems, which planned to incorporate more advanced algorithms for anomaly detection as well as new algorithms based on DL. Yan and Qu [26] introduced stacked denoising auto-encoders to solve the aero-engine sensor fault diagnosis problem. Miao *et al.* [27] established a method based on deep long short-term memory (LSTM) networks for joint learning of degradation assessment and remaining useful life prediction of aero-engines. Wu *et al.* [28] put forward a method based on the deep belief network (DBN) for aero-engine condition monitoring. The demonstration was carried out on the maintenance data of CF6-80C2A5 civil aircraft engines. Che *et al.* [29] proposed a DBN-based aero-engine fault fusion diagnosis model and validated its effectiveness by the simulation data of the Pratt & Whitney JT9D-74R engine. Liu and Li [30] provided a brief introduction to the research progress made by the United Technologies Corporation (UTC) in the use of DBN for aero-engine fault prediction. As far as we know, there has been no published literature on the DL application to the health monitoring of LREs.

One-class classification (OCC) algorithms are commonly used in areas such as outlier/novelty detection and concept learning, where the negative class is either absent, poorly sampled or not well defined [31]. Among them, methods based on one-class support vector machine (OCSVM) are noticeable and have shown great potential in fault detection. Sarmiento *et al.* [32] proposed a method based on OCSVM and principal component analysis (PCA) to detect faults in reactive ion etching systems through optical emission spectroscopy data. Shin *et al.* [33] applied OCSVM to fault detection and classification in electro-mechanical machinery from vibration measurements and compared it with multilayer perception (MLP). Results showed that the performance of OCSVM with appropriate parameters was superior to that of MLP. Mahadevan and Shah [34] realized fault detection based on OCSVM and demonstrated its effectiveness on the benchmark Tennessee Eastman problem and an industrial real-time semiconductor etch process dataset. Huang *et al.* [35] provided a method for mechanical fault detection of high voltage circuit breakers based on OCSVM and wavelet time-frequency entropy (WTFE). Anaissi *et al.* [36] adopted an OCSVM method with an automated parameter tuning algorithm for damage detection and demonstrated its effectiveness by the successful implementation in sensing data collected from the Sydney Harbour Bridge and vehicle-mounted sensors. In these research work, the fault detection process is usually assisted by certain feature extraction or selection methods, such as PCA [32] and WTFE [35]. However, it often seems insufficient to base the feature extraction on artificial design and expertise when facing an extremely complex system whose physical and chemical processes are too intricate to be fully understood.

The contribution of our work is twofold. First, we propose a fault detection method for the steady-state process of LRE based on convolutional auto-encoder (CAE) and OCSVM. We use CAE to extract features from normal data automatically, and train the OCSVM as the fault detector by the extracted features. Then we integrate the well-trained CAE and OCSVM to establish the fault detection method. The effectiveness of the method is verified by the ground hot-fire test data of a large LRE. Second, we elaborate on the effect of sample sizes and domain knowledge on the fault detection performance and conduct the comparative study with other traditional methods.

The rest of the paper is organized as follows. We explain the problem background in section 2, which is a brief description of the LRE and the ground hot-fire test data. Section 3 presents the preliminary knowledge about CAE and OCSVM respectively. Section 4 gives details about the proposed fault detection method for LRE steady-state process. Section 5 reports the results and discussions. The conclusions and future works are finally summarized in section 6.

II. PROBLEM BACKGROUND

In this section, we give a brief introduction to the problem background. First, the basic structure and operating principles

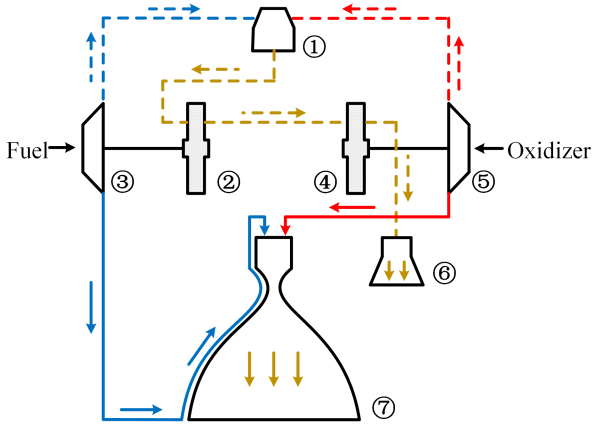


FIGURE 1. Main structures and working process of the LRE. (① Gas generator ② Fuel turbine ③ Fuel pump ④ Oxidizer turbine ⑤ Oxidizer pump ⑥ Turbine exhaust pipe ⑦ Thrust chamber).

TABLE 1. Information of data from the 10 ground hot-fire tests.

Operating condition	Test No.	Remarks
Normal	01, 02, 03, 04 05, 06, 07, 08	--
Fault	09	During the LRE test, most of the monitoring parameters plunged around 42.5s. The LRE was shut down at 43.2s and the whole test was not complete. The post-analysis results showed that the machining residuals in the pipeline blocked the venturi in the oxidizer subsystem, which led to the reduction of the oxidizer flow and the decrease of the operating condition.
	10	Due to the failure of the fuel turbine end seal, the rotating speed of the fuel turbine exceeded the lower limit during the test, leading to the automatic emergency shutdown.

TABLE 2. The selected parameters.

No.	Parameter (Unit)	Symbol	No.	Parameter (Unit)	Symbol
1	Inlet pressure of valve in front of fuel pump (MPa)	p_{ivfp}	12	Outlet pressure of fuel cooling jacket (MPa)	p_{ofcj}
2	Inlet pressure of valve in front of oxidizer pump (MPa)	p_{ivop}	13	Outlet pressure of oxidizer pump (MPa)	p_{oop}
3	Inlet pressure of oxidizer turbine (MPa)	p_{iot}	14	Outlet pressure of fuel pump (MPa)	p_{ofp}
4	Outlet pressure of fuel turbine (MPa)	p_{oft}	15	Inlet temperature of valve in front of fuel pump (K)	T_{ivfp}
5	Isolation chamber pressure of oxidizer turbine (MPa)	p_{icot}	16	Temperature after fuel pump (K)	T_{fp}
6	Inlet pressure of fuel turbine (MPa)	p_{ift}	17	Inlet temperature of valve in front of oxidizer pump (K)	T_{ivop}
7	Chamber pressure of gas generator (MPa)	p_{cgg}	18	Temperature after oxidizer pump (K)	T_{op}
8	Thrust chamber pressure (MPa)	p_{tc}	19	Fuel flow rate of LRE (kg/s)	$\dot{m}_f$
9	Fuel plenum pressure of gas generator (MPa)	p_{fpgg}	20	Oxidizer flow rate of LRE (kg/s)	$\dot{m}_o$
10	Oxidizer plenum pressure of gas generator (MPa)	p_{opgg}	21	Rotational speed of oxidizer turbine (rpm)	n_{ot}
11	Oxidizer plenum pressure of thrust chamber (MPa)	p_{optc}	22	Rotational speed of fuel turbine (rpm)	n_{ft}

of the LRE are described. Then an overview of the involved ground hot-fire test data is given.

A. DESCRIPTION OF LIQUID ROCKET ENGINE

Fig. 1 shows the main structures and working process of a large LRE with the gas generator cycle. The LRE mainly comprises a gas generator, a fuel turbine, a fuel pump, an oxidizer turbine, an oxidizer pump, a turbine exhaust pipe, and a thrust chamber. The related propellant feed system and control system are appropriately simplified to make the schematic diagram more concise.

During the normal operation, a small amount of the propellant is fed into the gas generator and burnt into high-temperature and high-pressure gas, driving both the fuel and oxidizer turbopump components to pressurize the propellant main feedlines. Then the gas is emitted into the atmosphere through the turbine exhaust pipe. Most of the propellant enters the thrust chamber to burn and generate thrust. Wherein, the fuel cools the thrust chamber via cooling channels in the chamber wall before entering the thrust chamber.

B. DATA DESCRIPTION

The working conditions of simple components, such as bearings, are more stable and the collected data usually have better repeatability. While for LRE, some design details may be improved during the whole test period, together with the complex structures and numerous monitoring parameters, leading to subtle differences and more uncertainties in working conditions among different tests, which increases the difficulty of model training. On the other hand, it may help us to obtain a model with better generalization performance, so that the model is robust to some normal fluctuations of working conditions within a certain range.

The data used here come from 10 ground hot-fire tests of a large LRE, including the steady-state process data of 8 normal tests and 2 fault tests with the sampling frequency of 50Hz. The operating conditions and 22 parameters of the LRE tests are detailed in the following Table 1 and Table 2.

III. PRELIMINARIES

In this section, we introduce some preliminary knowledge of CAE and OCSVM, which are the basic components of the proposed method.

A. CONVOLUTIONAL AUTO-ENCODER

CAE was first proposed by Masci *et al.* [37] in 2011 as an unsupervised feature learning method. It can be regarded as a special case of the traditional auto-encoder (AE) [38]. Traditional fully-connected AEs generally ignore the two-dimension structure in the data, which not only causes problems when dealing with real data but also introduces a large number of redundant parameters forcing each feature to be global. While CAE's weights are shared among all the positions of the whole input, just like the convolutional neural network (CNN), which ensures the spatial locality of the features and greatly reduces the number of parameters in the network. Retaining the spacial information of the two-dimension signals makes the extracted features more general.

According to reference [37], for a single channel input x, the encoding process for the k-th feature map can be expressed as:

$$h^k = \sigma(x * W^k + b^k) \tag{1}$$

where h is the encoding result, σ is an activation function, which means the rectified linear unit (ReLU) function here, W is a weight matrix, $*$ donates the convolution operation and b is the bias.

The decoding process can be expressed as:

$$y = \sigma(\sum_{k \in H} h^k * \tilde{W}^k + c) \tag{2}$$

where y is the decoding result, H is the set of latent feature maps, $\tilde{W}$ is the weight matrix, which can be obtained by the flip operation over both dimensions of the weights W or through training independently, and c is the bias.

The cost function to minimize is usually the mean squared error (MSE), which is calculated by the input x_i and its reconstruction y_i as following:

$$E(\theta) = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{3}$$

With the help of the backpropagation algorithm [39], the gradient of the error function with respect to the parameters can be obtained using the following formula:

$$\frac{\partial E(\theta)}{\partial W^k} = x * \delta h^k + \tilde{h}^k * \delta y \tag{4}$$

where δh and δy are the deltas of the hidden states and the reconstruction. Then the weights can be updated by stochastic gradient descent.

Thanks to the rapid development of various open-source frameworks for DL, researchers do not need to pay too much attention to the details of the algorithms and can implement the DL algorithms quickly and accurately. The construction and training of the CAE model here are completed with the help of TensorFlow [40], [41], which is one of the most popular open-source frameworks for ML and DL today.

B. ONE-CLASS SUPPORT VECTOR MACHINE

OCSVM was developed from the classical support vector machine (SVM) and first proposed by Schölkopf *et al.* [42] in 2001. As a supervised method, SVM is originally designed for binary classification problems, where both the well-prepared positive and negative samples are required to obtain the discriminating classifiers. In practice, counterexamples are either rare, entirely unavailable or statistically unrepresentative [43] in many cases, which motivates the development of OCC methods.

For LRE fault detection, there are many difficulties in data preparation, mainly involving the following three aspects. Firstly, there are so many different types of faults because of the complex structures and numerous components that it is almost impossible to guarantee that all kinds of faults have been included in one dataset. Secondly, due to the extreme working conditions, the LRE faults always develop rapidly and probably lead to catastrophic consequences. Therefore, researchers generally do not specially carry out fault experiments. Thirdly, even if a fault occurs during the test, effective measures will be taken immediately, usually, shutdown, to prevent the LRE from longstanding operation under fault conditions and even leading to disastrous consequences. For the above reasons, there are usually abundant data describing the normal operation process, while abnormal cases are rare. The inevitable data imbalance makes OCC methods suitable to solve LRE fault detection problem.

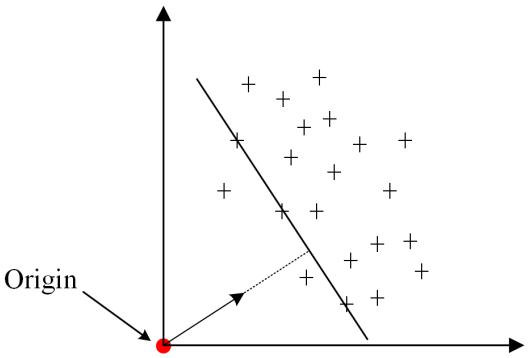


FIGURE 2. Principle diagram of OCSVM.

The basic principle of OCSVM is depicted in Fig. 2. Only normal data are used as training samples, which are mapped by the kernel function into high-dimensional space. The origin is treated as the only abnormal sample point, so that there is an optimal hyperplane that can separate the origin from the training samples and maximize the interval.

For a given dataset $Z = \{z_i\}_{i=1}^n, z_i \in R^m$, n is the size of the training set. For a kernel function $K(z_i, z_j) = \phi(z_i)^T \phi(z_j)$, where ϕ is the mapping function, the decision function is given by

$$f(z) = sgn(w \cdot \phi(z) - \rho) \tag{5}$$

where w is a perpendicular vector to the decision boundary and ρ is the bias term, which can be derived from the following optimization problem:

$$\begin{aligned} &\min_{w,\xi,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho \\ &\text{subject to } w \cdot \phi(z_i) \geq \rho - \xi_i, \quad \xi_i \geq 0 \end{aligned} \tag{6}$$

here ξ_i is a slack variable allowing points to violate the boundary constraints, and ν is a user-defined parameter to control the ratio of anomalies in the training set.

To solve this problem, multipliers $\alpha_i, \beta_i \geq 0$ are introduced to construct a Lagrangian function as follows [42]:

$$\begin{aligned} L(w, \xi, \rho, \alpha, \beta) = &\frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_i \xi_i - \rho \\ &- \sum_i \alpha_i (w \cdot \phi(z) - \rho + \xi_i) \\ &- \sum_i \beta_i \xi_i \end{aligned} \tag{7}$$

Set the partial derivatives of the primal variables to be zeros:

$$w = \sum_i \alpha_i \phi(z_i) \tag{8}$$

$$\alpha_i = \frac{1}{\nu n} - \beta_i \leq \frac{1}{\nu n}, \quad \sum_i \alpha_i = 1 \tag{9}$$

With the help of the kernel function, the dual form of the above optimization problem is as follows:

$$\min_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(z_i, z_j)$$

$$\text{subject to } 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_{i=1}^{n} \alpha_i = 1 \tag{10}$$

where $K(z_i, z_j)$ is usually the Gaussian kernel function, which is as follows:

$$K(z_i, z_j) = exp\left(-\frac{\|z - z_j\|^2}{2\sigma^2}\right) \tag{11}$$

where σ is the width parameter of the Gaussian kernel function.

For a new point z_{new}, the decision function is obtained by solving the dual problem:

$$f(z) = sgn\left(\sum_{i=1}^{n} \alpha_i K(z_i, z_{new}) - \rho\right) \tag{12}$$

where ρ can be calculated by the following formula:

$$\rho = \sum_i \alpha_i K(z_i, z_j) \tag{13}$$

More details can be found in the literature [42], [44].

IV. FAULT DETECTION METHOD FOR STEADY-STATE PROCESS OF LIQUID ROCKET ENGINE

In this section, we develop a fault detection method consisting of a CAE based feature extraction module and an OCSVM based fault detection module for the LRE steady-state process. In the following, we start by giving an outline of the whole detection method, and then detail the various aspects of it.

A. OUTLINE OF FAULT DETECTION METHOD

We propose a fault detection method based on CAE and OCSVM for the steady-state process of a certain type of large LRE. Here is the brief introduction of the method and the details are described in the following parts.

As shown in Fig. 3, the LRE hot-fire test data are collected by the data acquisition system and then split into samples after preprocessing. The samples are divided into the training set, validation set, and test set. The training process includes two stages. In the first stage, unsupervised learning is utilized to train the CAE for automatic data feature extraction. In the second stage, the features of training set extracted by well-trained CAE are fed into the OCSVM and semi-supervised learning is performed to obtain the one-class classifier. The validation set is mainly used for the hyperparameter selection in both training stages. Finally, the prepared CAE and OCSVM are connected to form the fault detection method. The test set is used to demonstrate the validity and evaluate the performance of the proposed fault detection method.

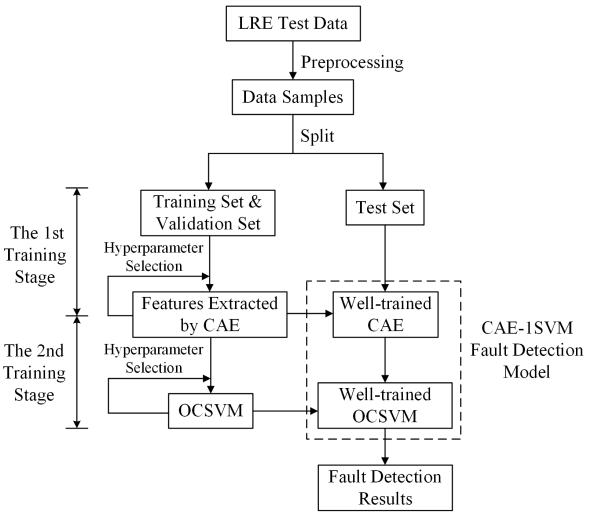


FIGURE 3. Flow diagram of the proposed fault detection method.

B. DATA PREPROCESSING

The LRE hot-fire test process can generally be divided into the startup phase, the steady-state phase and the shutdown phase. We only consider the steady-state phase here, so the data of the other two phases in the original data are removed. The data of the steady-state phase are preprocessed as follows.

1) SAMPLE SPLITTING METHOD

In the practical LRE test, the duration of each test is different, resulting in different data lengths. Besides, the data collected in one single test are too much to use directly. Therefore, the raw data need to be split into samples with a smaller size for further processing.

CAE aims to perform dimensionality reduction on data during feature extraction. So, the sample size is generally designed to be 2^a or a multiple of 2^a, which usually helps to improve the model performance. To explore the effect of the sample sizes on the method performance, we construct datasets with sample size 24×22, 48×22 and 96×22 respectively, where 24, 48 and 96 represent the number of consecutive sampling points in each sample, and 22 represents the number of parameters. Samples are obtained with the overlap strategy [45], as shown in Fig. 4. The shift is set to 1, which means that two contiguous samples differ by only one sampling point, which is 0.02s.

2) ZERO PADDING

As mentioned before, the size of the samples prefers 2^a or the multiple of 2^a. So, we add a column of zeros on each side of

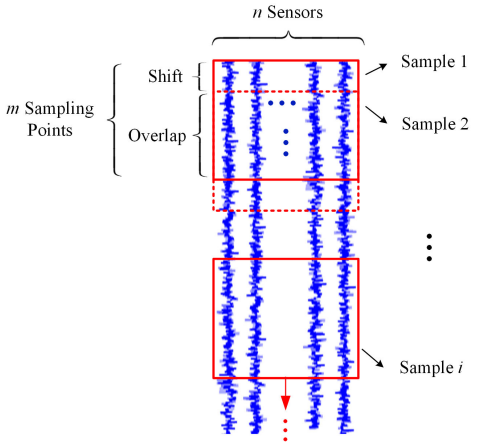


FIGURE 4. Schematic diagram of the sample splitting method with overlap strategy.

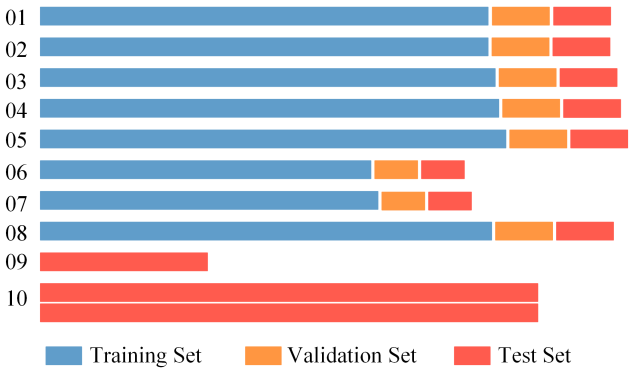


FIGURE 5. Distribution of samples in different datasets.

the sample, namely zero padding, which makes the sample size 24 × 24, 48 × 24 and 96 × 24 respectively.

Zero padding is a small trick frequently used in CNNs to adjust the sample to an appropriate size. Because the values of padding are all zeros, no noise is introduced into the original data.

3) NORMALIZATION

The sample data contain four types of parameters, including pressure, temperature, mass flow and rotating speed, with different orders of magnitude from 10^{-1} to 10^{4}. It is necessary to standardize the samples of the raw data. The z-score normalization [46] is adopted to ensure the same influence of different parameters on the feature extraction process and the reliability of the obtained features.

Z-score normalization, also called standardization, is a classic normalization method to transform the feature component x to x', which is a centered, scaled version of x with the same size, zero mean and unit variance.

$$x' = \frac{x - \mu}{\sigma} \tag{14}$$

where μ and σ are the mean and standard deviation of that feature respectively. For matrix, z-score normalization is computed using the mean and standard deviation along each column of the matrix.

It is worth noting that the mean and the standard deviation taken from the training set are used to preprocess all the samples in the training set, the validation set, and the test set, to avoid introducing information from the test set during the training process.

4) SPLITTING STRATEGY FOR DATASETS

For LRE fault detection, the continuity criterion is usually introduced to reduce the false alarm rate and improve the robustness of the algorithm, that is, a failure is considered to occur when the method alarms three times successively. The disorder of samples and the random division of datasets will make it difficult to judge whether a fault has occurred. Therefore, the last 10% of the continuous samples from the 8 normal LRE tests are divided into the test set along with the samples from 2 fault LRE tests. Another 10% of the continuous samples of the normal LRE tests are chosen as the validation set and the remaining samples of the normal LRE tests constitute the training set. As in Fig. 5, the bar diagram shows the distribution of samples from LRE tests in different datasets. The lengths of the bar represent the numbers of samples and three different colors represent the training set, the validation set, and the test set, respectively. The exact numbers of samples included in each dataset are listed in Table 3.

C. FEATURE EXTRACTION BY CONVOLUTIONAL AUTO-ENCODER

Take the sample size of 96 × 24 as an example, the CAE built here uses a typical symmetric structure consisting of a coding process with 5 convolutional layers and a decoding process with 5 deconvolutional layers, as shown in Fig. 6. During a sample flows through the CAE, the original sample with the size of 96 × 24 is converted to the automatically extracted feature with the size of 3 × 3 × 16 through 5 coding layers, and then restored to an output with the same size of input via 5 decoding layers. The MSE is used to measure the deviation between the input and the output. After the unsupervised learning process, a 3 × 3 × 16 feature matrix can be extracted automatically by the well-trained CAE, which can be further reshaped into a 144-dimension feature vector as the input of the OCSVM.

The specific structures of the CAEs adopted for different input sizes are listed in Table 4. Some adjustments are made to the first two coding layers and the last two decoding layers to adapt to different sizes, and the network structures of the

TABLE 3. Number of samples in different datasets.

	Training set	Validation set	Test set
Number of samples	143126	17908	32476

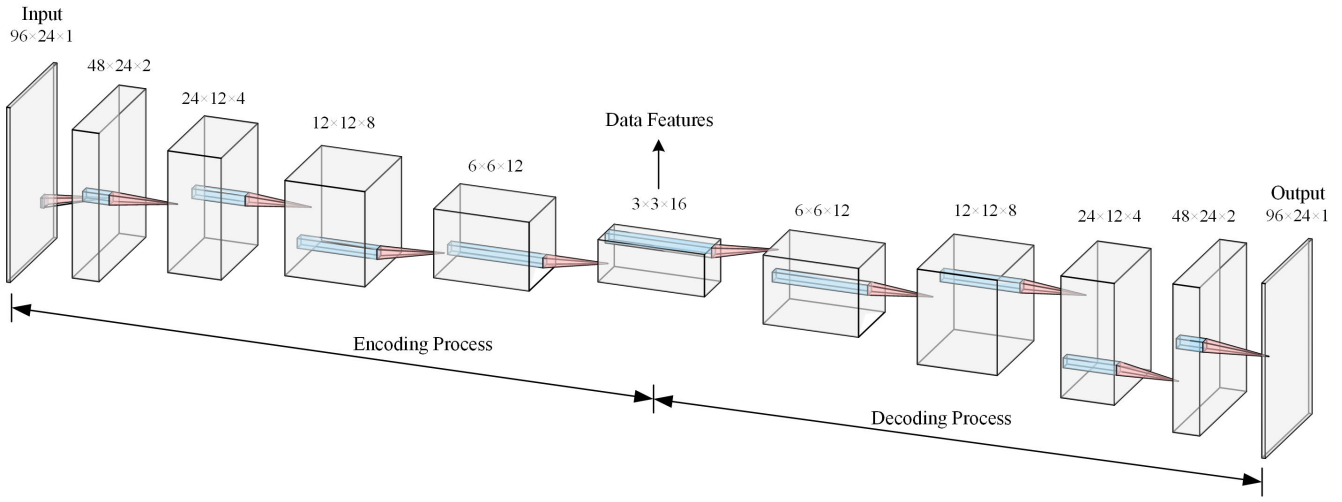


FIGURE 6. Architecture of the CAE with 96 × 24 input.

TABLE 4. Dimensions of each layer of the CAE with different input sizes.

Layers	Dimensions		
Input	24×24×1	48×24×1	96×24×1
Encode 1	24×24×2	48×24×2	48×24×2
Encode 2	12×12×4	24×12×4	24×12×4
Encode 3	12×12×8	12×12×8	12×12×8
Encode 4	6×6×12	6×6×12	6×6×12
Data features	3×3×16	3×3×16	3×3×16
Decode 1	6×6×12	6×6×12	6×6×12
Decode 2	12×12×8	12×12×8	12×12×8
Decode 3	12×12×4	24×12×4	24×12×4
Decode 4	24×24×2	48×24×2	48×24×2
Output	24×24×1	48×24×1	96×24×1

other parts are identical. The size of all convolution kernels is 3 × 3, and the activation functions are ReLU functions in all CAEs.

D. FAULT DETECTION BASED ON ONE-CLASS SUPPORT VECTOR MACHINE

The OCSVM with the Gaussian kernel function is trained by the features extracted by CAE from the training set, and the optimal hyperparameters are obtained with the help of the validation set and grid search method. Finally, the well-trained CAE and OCSVM are connected in series as the steady-state process fault detection method. The detection result is normal (no fault occurs) with the output +1, and an alarm is raised with the output −1. It is considered that a fault does occur when the detection method sends out alarms 3 times successively.

E. EVALUATION INDICATORS OF METHOD PERFORMANCE

For fault detection and diagnosis of simple components, as well as traditional image classification, the labels of the samples are usually known and definite. Even the samples in the test set have definite labels. Therefore, the test data can be used to quantitatively evaluate the performance of the detection or classification methods by some general indicators, such as precision, recall and F-measure [47]. For example, motor bearings were seeded with faults using electro-discharge machining (EDM) in the most widely used seeded fault test data provided by the Bearing Data Center of the Case Western Reserve University. However, the situation may be much more complicated in practical applications.

For LRE fault detection, the most important task is to detect the anomalies and faults occurring during the operation process as early as possible without false alarms. The redline system (RS), whose basic principle is to monitor whether some important operating parameters exceed the pre-determined limits or thresholds [8], is usually used as a benchmark method to judge the performance of the proposed method. It is generally believed that the LRE has failed when the RS issues an alarm. However, most of the faults do not occur suddenly. In general, the fault symptoms probably have appeared before the monitoring parameters exceed the thresholds. But we cannot accurately confirm from which moment the LRE has started to behave abnormally. Especially in the early stage of some soft faults, it may be difficult to distinguish the fault symptoms from the performance fluctuations and environmental noise. In this case, the labels of samples that come from the hot-fire test with fault are unidentified, especially for the samples from the period before the RS alarms. So, the mentioned commonly used evaluation indicators are inapplicable here.

Therefore, all the samples from the hot-fire tests occurring faults are divided into the test set. Meanwhile, about 10% of the samples from the normal tests are put into the test set. Two evaluation indicators are involved here. The qualitative indicator is whether a false alarm is issued on the normal data, and the quantitative one is the alarm time, which is certainly the earlier the better, on the data from the hot-fire tests with fault. The ideal result is that the proposed method does not cause false alarms on the normal data, and can detect faults on the data of fault tests as early as possible.

V. RESULTS AND DISCUSSIONS

To demonstrate its effectiveness, we compare the fault detection results of the proposed method with that of the redline system (RS), adaptive threshold algorithm (ATA) [48] and back-propagation neural network (BPNN) [49]. We also investigate the effect of sample sizes and domain knowledge on the performance of the proposed method. All experiments are run on NVIDIA GeForce GTX 750Ti 2 GB and the Microsoft Windows 7 operating system. The programming language is Python 3.5. The deep learning library is TensorFlow-gpu 1.4 and the machine learning library is scikit-learn 0.20.3.

A. MODEL TRAINING

The training process is critical and directly related to the performance of the method. Therefore, it is necessary to further detail the training process to make the proposed method clearer.

1) CAE TRAINING

The weights are initialized by the Xavier algorithm [50] which can help to keep the activation values and back-propagated gradient values in a reasonable range. Minibatch gradient descent [51] and Adam [52] are applied comprehensively to minimize the reconstruction error, namely the MSE. In minibatch gradient descent, the error with respect to some subset of the total training set is computed and used in the backpropagation, which strikes a balance between the efficiency and robustness of the algorithm. While vanilla mini-batch gradient descent is susceptible to saddle points and ill-conditioning. Adam, a first-order gradient-based optimization algorithm, can effectively overcome the mentioned shortcomings and has the advantages of high computational efficiency, little memory requirements and invariance to gradient diagonal rescaling. L^2 regularization [53], the most common used regularization method, is applied to improve the generalization performance of the model by encouraging the sum of the squares of the parameters to be small. The hyperparameters, such as initial learning rate and regularization rate, are selected with the help of the random search strategy [54] and determined when the set of hyperparameters minimizes the loss on the validation set.

Take the dataset with the sample size of 24 $\times$ 24 as an example, the loss curves of the CAE training process with optimal hyperparameters are depicted in Fig. 7. It can be seen that the losses on both the training set and validation set drop rapidly with the increase of training epochs in the early stage of the training process. After 1000 epochs of training, the losses are gradually stable and the training process can be stopped.

2) OCSVM TRAINING

OCSVM is established and trained with the help of the OneClassSVM function in scikit-learn toolkit [55]. The grid search strategy is used for the optimization of two critical hyperparameters, namely the *gamma* and *nu*. The former, *gamma*, is a parameter in the Gaussian kernel function, which determines the distribution of data after mapping to a new feature space. It also affects the number of support vectors and further affects the speed of training and prediction. *nu* indicates the upper bound on the fraction of training errors and a lower bound of the fraction of support vectors.

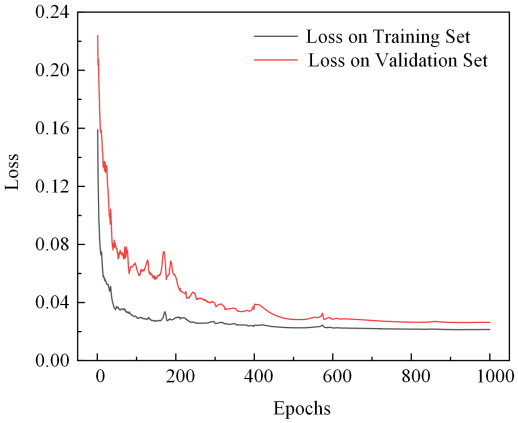


FIGURE 7. Loss curves of the CAE training process.

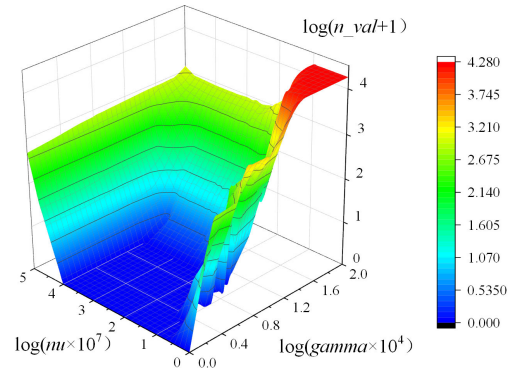
Fig. 8 shows the variation of training results with the two hyperparameters, *gamma* and *nu*. Logarithmic transformation is done to the hyperparameters and the training results, namely the number of false negatives, to make the trends clearer. As can be seen in Fig. 8(a) and (b), OCSVM can get better results on the validation set within a relatively wide range of *nu* when *gamma* is small. While in the ideal situation, we expect that OCSVM can perform well on both the validation set and the training set. In combination with Fig. 8(c) and (d), the choices of hyperparameters are limited in a narrow range when the number of false negatives is required to be small enough on the training set as well. In fact, the hyperparameters are probably can be selected from the yellow dashed area in Fig. 8(d). Further, we prefer to choose a hyperparameter combination with a larger *gamma* which usually means better generalization performance. The hyperparameter selections for all experiments in this paper follow such procedures and guidelines.

B. EXPERIMENT RESULTS

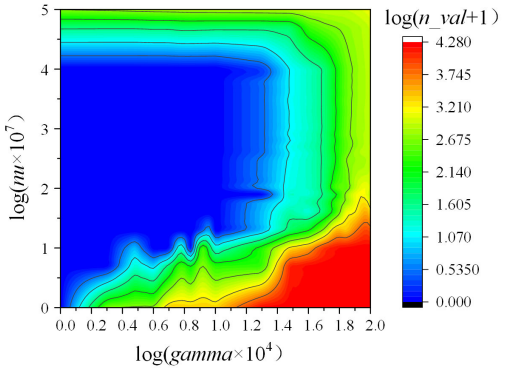
In this section, two sets of experiments are designed and discussed to study the influence of sample sizes and domain knowledge on fault detection results.

1) EFFECT OF SAMPLE SIZES

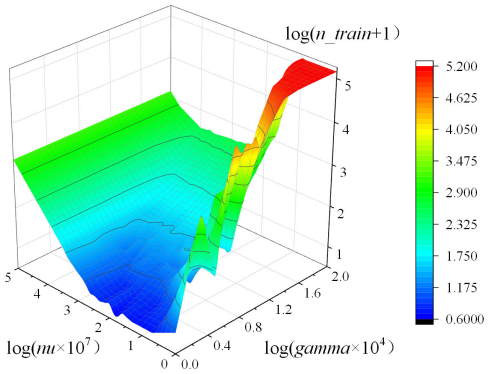
Since the number of monitoring parameters are fixed, the sample size here refers to the sample length. Three cases with sample lengths of 24, 48 and 96 are discussed, which means that 24, 48 and 96 consecutive sampling points are included in the samples of each case. Considering that the acquisition frequency is 50Hz, samples in three cases contain information collected within 0.48s, 0.96s and 1.92s respec-



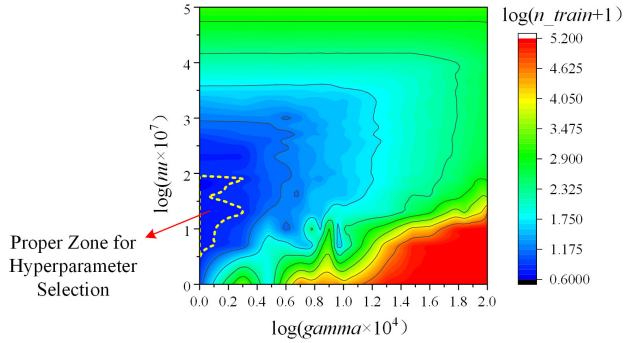
(a) Number of false negatives in the validation set.



(b) Number contour of false negatives in the validation set.



(c) Number of false negatives in the training set.



(d) Number contour of false negatives in the training set.

FIGURE 8. Training results with different hyperparameters.

tively. Intuitively, a larger sample size means that a single sample contains more information, which should lead to better detection performance.

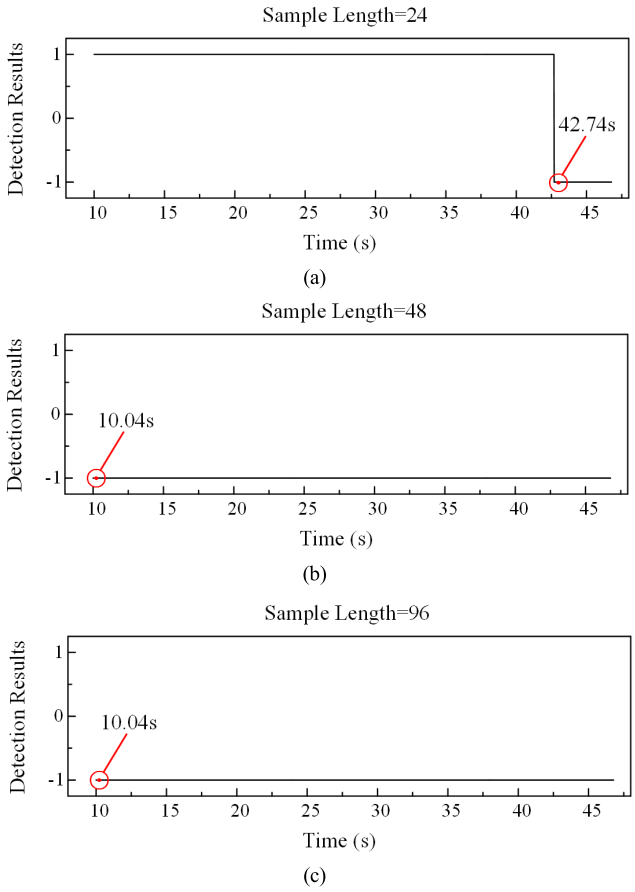


FIGURE 9. Detection results of Test09.

The sample to be detected is fed into the model and the detection results can be directly output. According to the continuity criterion, three consecutive output of −1 indicates that the system is in a failure mode. Results show that the method can achieve no false alarms on normal data by reasonably selecting the hyperparameters. The following mainly discusses the detection results of the 2 fault test runs.

The detection results are shown in Fig. 9 and Fig. 10. For Test09, the detection method alarms at 42.74s when the sample length is 24. As the sample length increases, all the samples from the steady process of Test 09 are judged as -1, which can be seen in Fig. 9(b) and (c). It means that the fault is detected at the very beginning of the LRE steady-state process. From Table 1, we know that the fault was caused by the machining residual in the pipeline, which was probably produced during the stage of machining or assembling. So, the residual was probably to have been in the pipeline from the beginning, not generated in the middle of the test run. At the start of the hot-fire test, the dispersed residual could only have a limited impact on certain parameters. However, as the test ran, the residual might flow with the propellant and gather together at specific locations (here, the venturi), causing more serious problems (here, the venturi was blocked and the working conditions declined rapidly). For Test10,

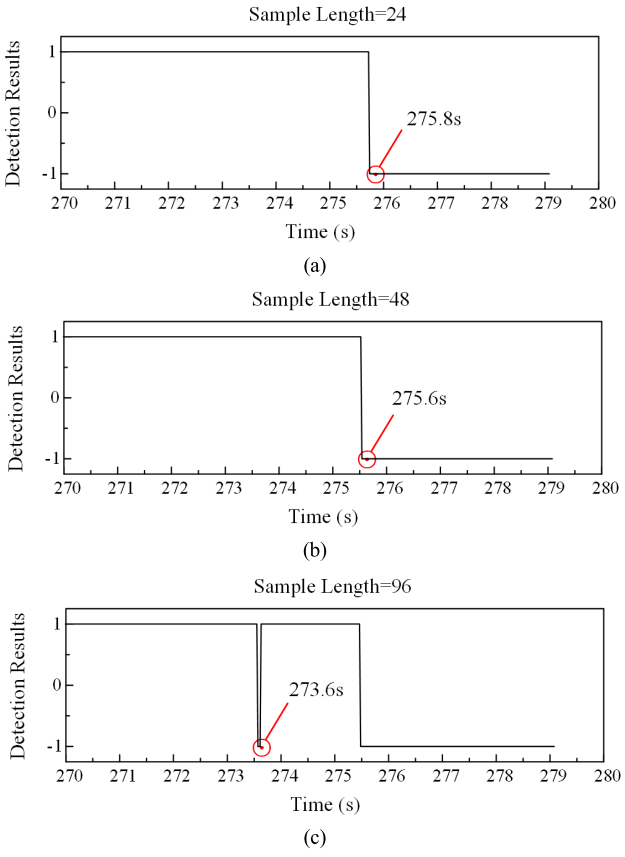


FIGURE 10. Detection results of Test10.

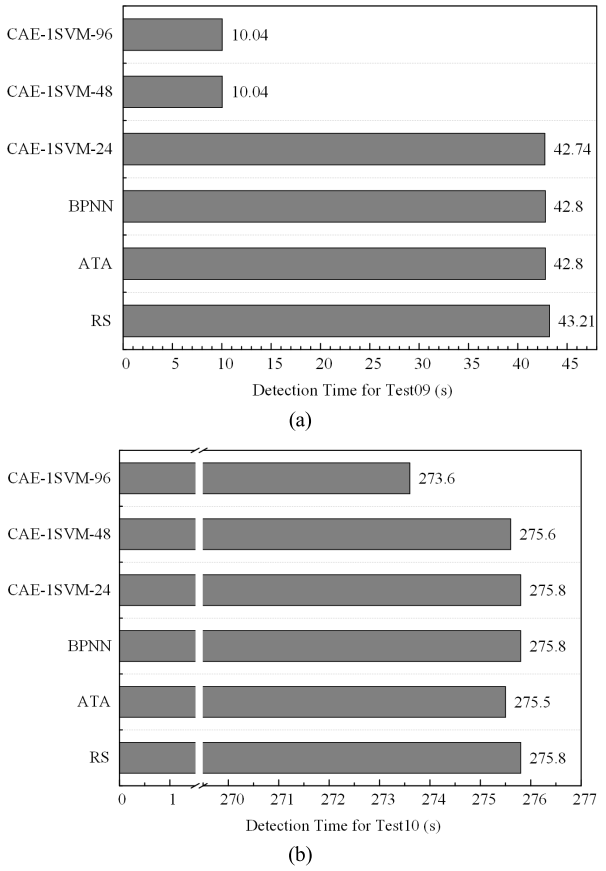


FIGURE 11. Comparison of fault detection results.

the longer the sample length is, the earlier the fault is detected. Especially when the sample length is 96, the detection time is advanced about 2 seconds, which is a great improvement for LRE fault detection. It is worth noting that, as can be seen from Fig. 10(c), the outputs become +1 after the 273.6s alarm, and the detection method alarms again at 275.52s. This indicates that the incipient fault has a limited influence on the system performance and some fluctuations may be caused on some parameters around the normal range. However, as the fault develops, the fault characteristics become more and more obvious, and the detection results are more definite and stable.

In general, as the sample size increases, the detection method greatly enhances the fault detection capability during the test run.

When compared with the previous methods, such as RS, ATA, and BPNN, the proposed method can achieve at least comparable results, as shown in Table 5 and Fig. 11. With the increase of the sample length, the advantages of the proposed method gradually emerge, and there is a significant advance in the fault detection time.

2) EFFECT OF DOMAIN KNOWLEDGE

We also study the influence of domain knowledge on the performance of the detection method. Domain knowledge here refers to some simple knowledge about the LRE structure. Specifically, the data collected by different sensors is rearranged according to the relationship between the monitoring parameters and LRE structures. Generally, the degree of coupling among parameters is higher when these parameters have closer relationships in structure. So, the general principle is to put parameters belonging to the same components or subsystems together according to the LRE structural composition, namely the domain knowledge.

TABLE 5. Comparison of fault detection results.

Test No.	RS	ATA	BPNN	CAE-1SVM with different sample lengths		
				24	48	96
01	43.21	42.8	42.8	42.74	10.04	10.04
02	275.80	275.5	275.8	275.8	275.6	273.6

The process of parameter rearrangement mainly follows three principles. Firstly, the parameters belonging to the same components are put together, as shown in Fig. 12. Secondly, the parameters collected along the fuel or oxidizer flow path are put together. Thirdly, the parameter order is consistent with the structural and logical relationships among the parameters. For example, the inlet parameters of the same components are arranged in front of the outlet parameters, the turbine parameters are arranged in front of the pump

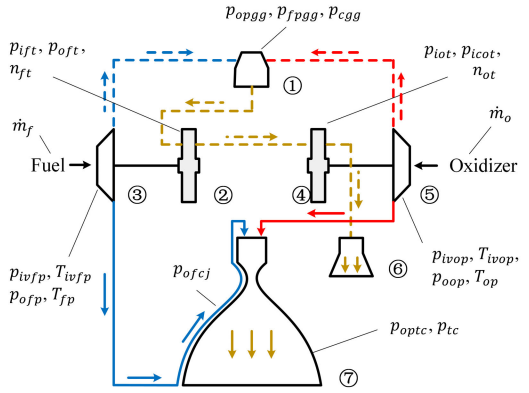


FIGURE 12. Parameters collected from different components.

TABLE 6. Rearrangement results of 22 parameters.

Initial Order		Rearrangement Results	
No.	Parameter	No.	Parameter
1	p_{ivfp}	1	p_{opgg}
2	p_{ivop}	2	p_{fpgg}
3	p_{iot}	3	p_{cgg}
4	p_{oft}	4	p_{ift}
5	p_{icot}	5	p_{oft}
6	p_{ift}	6	n_{ft}
7	p_{cgg}	7	p_{ivfp}
8	p_{tc}	8	T_{ivfp}
9	p_{fpgg}	9	p_{oft}
10	p_{opgg}	10	T_{fp}
11	p_{optc}	11	p_{ofcj}
12	p_{ofcj}	12	$\dot{m}_f$
13	p_{oop}	13	p_{iot}
14	p_{ofp}	14	p_{icot}
15	T_{ivfp}	15	n_{ot}
16	T_{fp}	16	p_{ivop}
17	T_{ivop}	17	T_{ivop}
18	T_{op}	18	p_{oop}
19	$\dot{m}_f$	19	T_{op}
20	$\dot{m}_o$	20	$\dot{m}_o$
21	n_{ot}	21	p_{optc}
22	n_{ft}	22	p_{tc}

parameters, and so on. The final rearrangement results are listed in Table 6.

The detection results are illustrated in Fig. 13 and 14, and the effect of domain knowledge is depicted in Fig. 15. For Test09, the detection method sends an alarm at 42.64s when the sample length is 24, which is slightly improved by 0.1s than that without domain knowledge. When the sample length increases, the detection method outputs -1 during the whole detection process. For Test10, when the sample length is 24 and 48, the fault is detected at 275.64s and 275.58s respectively, which shows slight improvement compared to the cases without domain knowledge. The fault is detected for the first time at 162.68s with sample length 96. It is more than 100s ahead of that without domain knowledge (273.6s), which can be considered as a significant improvement. During the entire detection process, it is a remarkable fact that

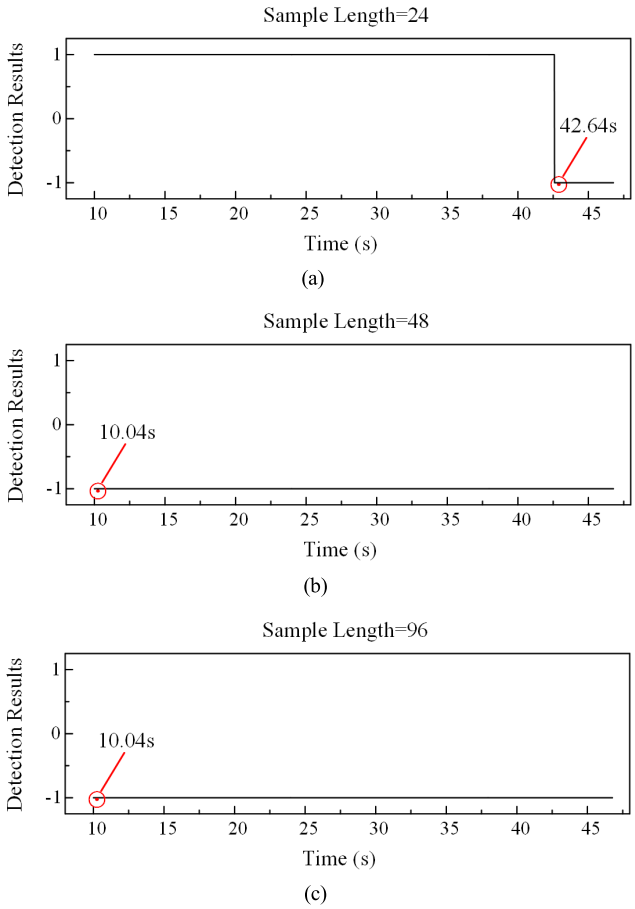


FIGURE 13. Detection results of Test09 with domain knowledge.

the detection results transition repeatedly between fault and normal states, and the detection results are stable to the fault state until 273.38s. A similar situation occurs without domain knowledge, but not so obvious, as indicated in Fig. 10(c). It can be seen that the introduction of domain knowledge significantly improves the sensitivity of the proposed method to incipient faults.

C. DISCUSSIONS

The proposed CAE-1SVM method can effectively realize the fault detection of LRE, and can meet the timeliness requirement. Compared with the traditional RS, ATA, and BPNN, the new method can achieve at least comparable or, in most cases, better results.

Samples with longer length contain more information about the LRE operation status. The introduction of domain knowledge also enriches the information content. Both contribute to the improvement of the detection performance. In addition to the advancement of detection time, several other advantages need further explanation.

1) Better anti-noise performance. The proposed method is more concise in data preprocessing, and no special measures are needed to denoise the original data.

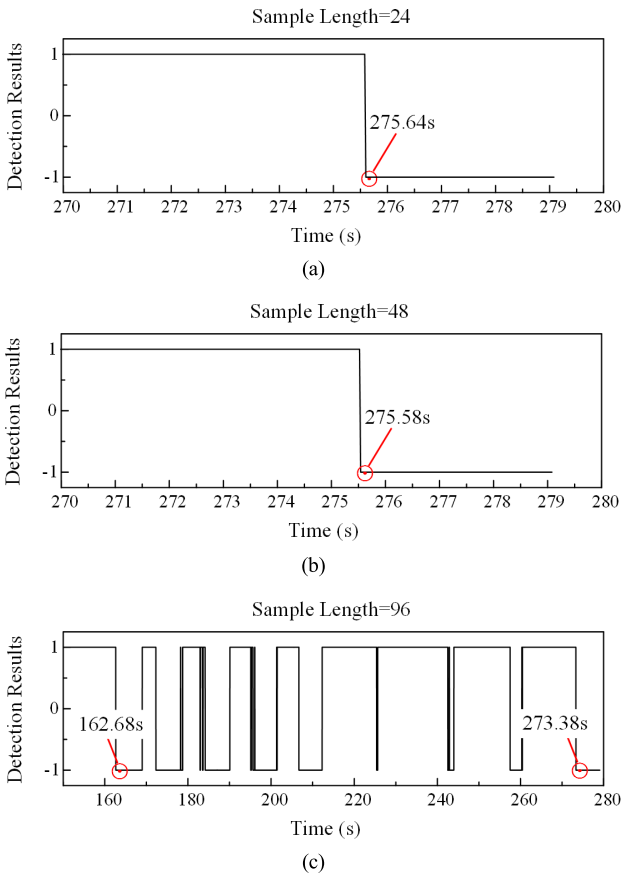


FIGURE 14. Detection results of Test10 with domain knowledge.

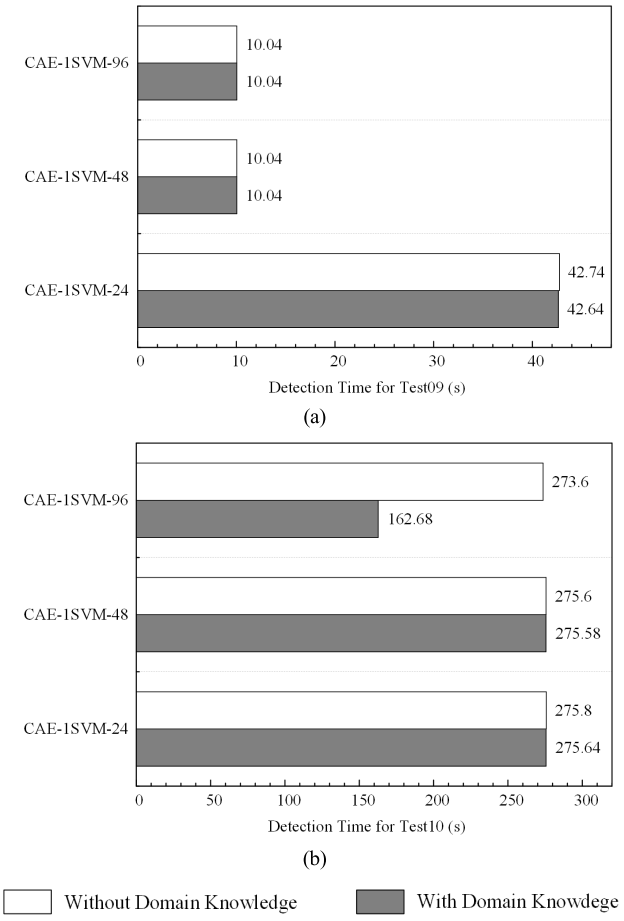


FIGURE 15. The Impact of domain knowledge on detection performance.

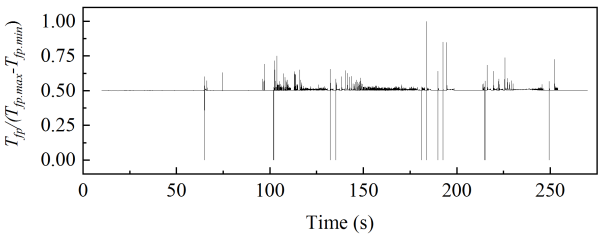


FIGURE 16. Temperature after fuel pump in Test04.

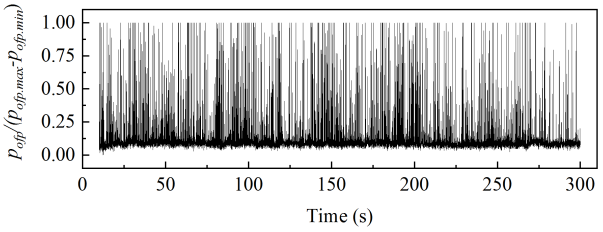


FIGURE 17. Outlet pressure of fuel pump in Test07.

The LRE working conditions are usually extreme and complex so that a large amount of noise is inevitably introduced to the monitoring system. Two normalized parameter fragments are taken from the original data of Test04 and Test07 respectively and depicted in Fig. 16 and Fig. 17. A lot of noise exists in the data, and occasionally there are even some invalid data points with a value of 0 as shown. The traditional approaches usually use filters to denoise data, or some artificially designed algorithms to remove coarse errors in the data [49]. Mainly thanks to the CAE based automatic feature extraction method, the proposed method is more robust to noise and accidental errors, so that it is not necessary to take special denoising measures for the original data.

2) Better generalization performance. As mentioned, the working conditions of any two hot-fire tests cannot be exactly the same, which leads to higher requirements on the generalization performance of the fault detection method. The results show that the proposed CAE-1SVM method can effectively extract common features from normal data of multiple test runs and realize the fault detection of new test runs. However, previous methods, such as the mentioned RS, ATA, and BPNN, generally come down to the threshold judgment problem. It is usually necessary to make appropriate adjustments to some key parameters, such as the position and the bandwidth of the thresholds, to meet the

timeliness of fault detection and avoid false alarms when facing new working conditions, which weakens the generalization performance of these methods to some extent.

VI. CONCLUSION AND FUTURE WORKS

This paper proposes a fault detection method combining convolution auto-encoder and one-class support vector machine for liquid rocket engine steady-state process. The automatic feature extraction is performed by the convolution auto-encoder. The fault detection is realized by the one-class support vector machine fed by the extracted features. To the best of our knowledge, this is the first time to apply deep learning methods in the field of fault detection for large liquid rocket engines. The effectiveness of the proposed method is validated by the real data collected from the ground hot-fire tests of the large liquid rocket engine, and its performance is compared with that of the redline system, adaptive threshold algorithm, and back-propagation neural network. Results show that the proposed method can effectively detect faults without false alarms and meet the timeliness requirements. The effect of sample sizes and domain knowledge on the performance of the proposed method is also discussed. The results suggest that both the longer sample length and the introduction of domain knowledge can help improve the performance of liquid rocket engine fault detection, especially for the incipient fault detection. Some other advantages of the proposed method are also discussed further.

This work is only a preliminary attempt for the application of deep learning in the field of liquid rocket engine fault detection. There is still much room for improvement.

1) The network scale of the deep learning model is much larger than the traditional machine learning methods because of more training parameters and hyperparameters. Therefore, higher requirements on hardware conditions and computing methods are put forward for the method training and operation. With the rapid development of related theories, methods, and hardware for deep learning, these issues are expected to be resolved in the near future, especially for the liquid rocket engine ground tests and post off-line analysis.
2) The convolutional auto-encoder and one-class support vector machine in the proposed method are training respectively, which may lead to the mismatch between the feature extraction process and the need for fault detection. The advantages of deep learning methods are not fully utilized. The joint optimization of the feature extraction and the fault detection process is also a valuable research direction.
3) The proposed method cannot start to function until the liquid rocket engine runs into the steady-state process for a short period of time, because the input of the method needs data collected within a certain period of time. It results in a short gap period at the beginning of the steady-state stage of the LRE operation. The combined utilization of traditional and DL based methods can effectively alleviate this problem.

In summary, the proposed fault detection method for LRE steady-state process is effective. Meanwhile, we believe that this is just the beginning and there is still a lot of room for improvement, which is worth further exploration.

ACKNOWLEDGMENT

The authors gratefully acknowledge A. LeNail [56] for his work in developing NN-SVG which helps them to create the neural network architecture drawing in Fig. 6, and the anonymous referees for the valuable suggestions on this article.

REFERENCES

[1] K. Lee, J. Cha, S. Ko, S.-Y. Park, and E. Jung, "Fault detection and diagnosis algorithms for an open-cycle liquid propellant rocket engine using the Kalman filter and fault factor methods," *Acta Astronaut.*, vol. 150, pp. 15–27, Sep. 2018.
[2] J. Thomas Bellows, R. E. Brewster, and E. Bekir, "OTV liquid rocket engine control and health monitoring," in *Proc. 20th AIAA/SAE/ASME Joint Propuls. Conf.*, Cincinnati, OH, USA, Jun. 1984.
[3] M. W. Hawman, "Health monitoring system for the SSME—program overview," in *Proc. 26th AIAA/SAE/ASME/ASEE Joint Propuls. Conf.*, Orlando, FL, USA, Jul. 1990.
[4] K. Gao, "Inductive monitoring system based fault detection for liquid-propellant rocket engines," in *Proc. CCC*, Hangzhou, China, 2015, pp. 6131–6135.
[5] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. de Oliveira, R. E. Vásquez, "A review on data-driven fault severity assessment in rolling bearings," *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196, Jan. 2018.
[6] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques-part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.
[7] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.
[8] J. Wu, "Liquid-propellant rocket engines health-monitoring—A survey," *Acta Astronaut.*, vol. 56, pp. 347–356, Feb. 2005.
[9] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis part III: Process history based methods," *Comput. Chem. Eng.*, vol. 27, pp. 327–346, Mar. 2003.
[10] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 45, pp. 144–156, Jul. 2018.
[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[12] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[15] P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," *Pattern Recognit. Image Anal.*, vol. 26, no. 1, pp. 9–15, 2016.
[16] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[17] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.
[18] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12 pp. 2493–2537, Aug. 2011.

[19] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
[20] G. Zhao, G. Zhang, Q. Ge, and X. Liu, "Research advances in fault diagnosis and prognostic based on deep learning," in *Proc. PHM*, Chengdu, China, 2016.
[21] Y.-J. Duan, Y.-S. Lv, J. Zhang, X.-L. Zhao, and F.-Y. Wang, "Deep learning for control: The state of the art and prospects," *Acta Automatica Sinica*, vol. 42, no. 5, pp. 643–654, 2016.
[22] H. Ren, J. F. Qu, Y. Chai, Q. Tang, and X. Ye, "Deep learning for fault diagnosis: The state of the art and challenge," (in Chinese), *Kongzhi yu Juece/Control Decis.*, vol. 32, no. 8, pp. 1345–1358, Aug. 2017.
[23] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018.
[24] X. Zhu, Z. Cai, J. Wu, Y. Cheng, and Q. Huang, "Convolutional neural network based combustion mode classification for condition monitoring in the supersonic combustor," *Acta Astronaut.*, vol. 159, pp. 349–357, Jun. 2019.
[25] E. A. Cortes and L. Rabelo, "An architecture for monitoring and anomaly detection for space systems," *SAE Int. J. Aerosp.*, vol. 6, no. 1, pp. 81–86, 2013.
[26] B. Yan and W. Qu, "Aero-engine sensor fault diagnosis based on stacked denoising autoencoders," in *Proc. CCC*, Chengdu, China, 2016, pp. 6542–6546.
[27] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.
[28] Y. Wu, J. Li, X. Wu, and H. Wang, "Civil aviation engine health condition monitoring based on DBN deep learning theory," (in Chinese), *Comput. Meas. & Control*, vol. 25, no. 7, pp. 28–31, 2017.
[29] C. Che, H. Wang, X. Ni, and J. Hong, "Fault fusion diagnosis of aircraft engine based on deep learning," (in Chinese), *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 44, no. 3, pp. 621–628, 2018.
[30] F. Liu and Q. Li, "Development and application of big data technology in aviation industry," (in Chinese), *Telecommun. Eng.*, vol. 57, no. 7, pp. 849–854, Jul. 2017.
[31] S. S. Khan and M. G. Madden, "One-class classification: Taxonomy of study and review of techniques," *Knowl. Eng. Rev.*, Jan. 2014.
[32] T. Sarmiento, S. J. Hong, and G. S. May, "Fault detection in reactive ion etching systems using one-class support vector machines," in *Proc. ASMC*, 2005, pp. 140–143.
[33] H. J. Shin, D. H. Eom, and S. S. Kim, "One-class support vector machines—An application in machine fault detection and classification," *Comput. Ind. Eng.*, vol. 48, pp. 395–408, 2005.
[34] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *J. Process Control*, vol. 19, no. 10, pp. 1627–1639, 2009.
[35] N. Huang, H. Chen, S. Zhang, G. Cai, W. Li, D. Xu, and L. Fang, "Mechanical fault diagnosis of high voltage circuit breakers based on wavelet time-frequency entropy and one-class support vector machine," *Entropy*, vol. 18, no. 7, 2016.
[36] A. Anaissi, N. L. D. Khoa, and Y. Wang, "Automated parameter tuning in one-class support vector machine: An application for damage detection," *Int. J. Data Sci. Anal.*, vol. 6, no. 4, pp. 311–325, 2018.
[37] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. ICANN*, Espoo, Finland, 2011, pp. 52–59.
[38] F. N. Yuan, L. Zhang, J. T. Shi, X. Xia, and G. Li, "Theories and Applications of Auto-Encoder Neural Networks: A Literature Survey," (in Chinese), *Jisuanji Xuebao/Chin. J. Comput.*, vol. 42, no. 1, pp. 203–230, 2019.
[39] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.
[40] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467
[41] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," 2016, *arXiv:1605.08695*. [Online]. Available: https://arxiv.org/abs/1605.08695
[42] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
[43] S. D. Villalba and P. Cunningham, "An evaluation of dimension reduction techniques for one-class classification," *Artif. Intell. Rev.*, vol. 27, no. 4, pp. 273–294, 2007.
[44] A. Bounsiar and M. G. Madden, "One-Class Support Vector Machines Revisited," in *Proc. ICISA*, 2014.
[45] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
[46] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognit. Lett.*, vol. 22, no. 5, pp. 563–582, 2001.
[47] G. Hoang, A. Bouzerdoum, and S. Lam, "Learning Pattern Classification Tasks with Imbalanced Data Sets," in *Pattern Recognition*, P. Y. Yin, Ed. Rijeka, Croatia: InTech, 2009, pp. 193–208.
[48] T. Xie, H. Liu, W. Ding, and J. Wu, "Implementation of fault detection and alarm system based on ATA algorithm in liquid rocket engine," (in Chinese), *J. Rocket Propuls.*, vol. 31, no. 6, pp. 19–22, 2005.
[49] Q. Huang, J. Wu, H. Liu, and T. Xie, "Implementation of real-time fault detection algorithms based on neural network for liquid propellant rocket engines," (in Chinese), *J. Nat. Univ. Defense Technol.*, vol. 29, no. 5, pp. 10–13, 2007.
[50] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, May 2010.
[51] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016, *arXiv:1609.04747*. [Online]. Available: https://arxiv.org/abs/1609.04747
[52] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *Proc. ICLR*, 2015, [Online]. Available: https://arxiv.org/abs/1412.6980
[53] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proc. ICML*, Banff, AB, Canada, 2004, pp. 615–622.
[54] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.
[55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[56] A. LeNail, "NN-SVG: Publication-ready neural network architecture schematics," *J. Open Source Softw.*, vol. 4, no. 33, p. 747, 2019.

XIAOBIN ZHU was born in 1990. He received the B.S. degree in flight vehicle propulsion engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2013, and the M.S. degrees in aeronautical and astronautical science and technology from Air Force Engineering University, Xi'an, China. He is currently pursuing the Ph.D. degree with the College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China. His main research interests are dynamics, control, and health management of the propulsion systems.

YUQIANG CHENG received the B.S. degree in engineering mechanics from Zhejiang University, Hangzhou, China, in 2002, the M.S. and Ph.D. degree in aerospace propulsion theory and engineering from the National University of Defense Technology, Changsha, China, in 2004 and 2009, respectively. He is currently an Associate Researcher and the M.S. Supervisor with the College of Aerospace Science and Engineering, National University of Defense Technology. His main research interests include damage dynamics, damage-mitigating control, and health management for liquid rocket engines.

JIANJUN WU received the B.S. and M.S. degrees in thermal engineering from Tianjin University, Tianjin, China, in 1988 and 1991, respectively, and the Ph.D. degree from the National University of Defense Technology, in 1995. He is currently a Professor and the Ph.D. Supervisor with the College of Aerospace Science and Engineering, National University of Defense Technology. His research interests include spacecraft and propulsion system health management, advanced non-chemical space propulsion, and liquid rocket engine dynamics and control.

RUNSHENG HU received the B.S. degree in flight vehicle propulsion engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015, and the M.S. degree in aeronautical and astronautical science and technology from the National University of Defense Technology, Changsha, China, in 2017, where he is currently pursuing the Ph.D. degree. His main research interests are dynamic system simulation and control.

XING CUI received the B.S. degree in flight vehicle propulsion engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018. She is currently pursuing the master's degree with the College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China. Her main research interest is liquid rocket engine health management.