SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

January 2026, Vol. 69, Iss. 1, 112207:1–112207:16
https://doi.org/10.1007/s11432-024-4689-9

Reduced KICA reconstruction modeling approach with dual attributes for fault diagnosis

Meizhi LIU[1,2], Xiangyu KONG[1*], Changhua HU[1] & Huaifan LI[2]

[1]*School of Missile Engineering, PLA Rocket Force University of Engineering, Xi'an 710025, China*
[2]*School of Physics and Electronic Science, Shanxi Datong University, Datong 037009, China*

Received 19 December 2024/Revised 31 March 2025/Accepted 15 April 2025/Published online 4 January 2026

Abstract Fault reconstruction identifies fault types by seeking a particular fault subspace that can effectively eliminate the alarm signal. A reconstruction model based on kernel independent component analysis (KICA) is proposed to address the non-Gaussian and nonlinear characteristics in fault diagnosis. However, nonlinear faults are characterized by nonlinear subspaces, which typically exhibit high dimensionality, contributing to increased spatial requirements and computational complexity. Moreover, commonalities among different subspaces may cause the same alarm signal to be eliminated by multiple subspaces, introducing uncertainty into the diagnostic process. To address these issues, a reduced KICA reconstruction modeling approach with dual attributes (RD-KICA) for fault diagnosis is proposed in this paper. An undersampling method is investigated to construct a less numerous but informative training set, such that the dimensionality of extracted subspaces is significantly reduced. The ideas of fault reconstruction and pattern classification are incorporated within the same framework, allowing their advantages to be complementary. Furthermore, the fault magnitude is supplemented as another attribute and used to train Bayesian classifiers for further diagnosis. Finally, several experiments on a numerical simulation, Tennessee Eastman process (TEP), and a rocket servo system are performed to validate the efficiency and benefits of the proposed method.

Keywords fault diagnosis, fault subspace, fault magnitude, feature vector selection, kernel independent component analysis

Citation Liu M Z, Kong X Y, Hu C H, et al. Reduced KICA reconstruction modeling approach with dual attributes for fault diagnosis. Sci China Inf Sci, 2026, 69(1): 112207, https://doi.org/10.1007/s11432-024-4689-9

1 Introduction

Fault diagnosis is a key technology to ensure the safe and stable operation of complex systems. Recently, data-driven methods have gained significant attention due to their ability to dispense with the need for precise mathematical models and prior knowledge [1]. This approach leverages the vast amounts of process data generated by contemporary distributed systems in industrial processes, offering a promising and feasible means to extract valuable fault information [2].

Fault reconstruction is a typical data-driven approach to fault diagnosis, which originates from statistical process monitoring (SPM). The essence of this approach lies in the consensus that the historical fault data encapsulates the key fault characteristics. Through singular value decomposition (SVD) applied to labeled historical fault datasets, distinctive feature vectors characterizing different failure modes can be extracted, subsequently forming fault subspaces. It is noteworthy that within this framework, each fault category is mathematically represented by its corresponding fault subspace, where the precision of this representation directly dictates diagnostic reliability. During online diagnosis, these fault subspaces are utilized to reconstruct fault samples. If the alarm signal of a fault sample can be eliminated by a certain subspace, we can classify this fault sample as the type associated with this subspace. Over the past two decades, several advances [3–5] have been made in the study of fault reconstruction, but most of them are based on principal component analysis (PCA) or projection to latent structure (PLS). However, these methods primarily use first- and second-order information, insufficient for non-Gaussian processes. Furthermore, PCA- and PLS-based methods require the assumption of a multi-variate Gaussian distribution for estimating control limits, leading to potentially inaccurate limits for non-Gaussian processes [6].

In practice, process data often deviate from a strict Gaussian distribution. In such cases, ICA is a suitable alternative because it can effectively utilize the higher-order information in non-Gaussian data to extract mutually independent components. Moreover, there is a complex nonlinear relationship between the fault data and fault

* Corresponding author (email: xiangyukong@ldy.edu.rs)

Table 1 The advantages and disadvantages of fault reconstruction and pattern classification.

Method	Advantages	Disadvantages
Fault reconstruction	(1) Effective identification of unseen faults. (2) Better interpretability.	(1) Low diagnostic accuracy for historical faults. (2) Diagnostic accuracy depends heavily on the precise fault representation of the fault subspaces.
Pattern classification	(1) High diagnostic accuracy for historical faults. (2) Numerous optional classifier models.	(1) Unable to identify unseen faults. (2) Diagnostic accuracy depends on the completeness of the historical fault datasets. (3) Easy to fall into overfitting.

characteristics, so linear models may fail to capture these nonlinear characteristics [7]. Hence, kernel ICA (KICA) [8] was investigated as an effective method for nonlinear and non-Gaussian processes. Numerous derivations of KICA have been reported, including performance-relevant KICA [9], Fractal dimension-based dynamic kernel independent component regression [7], and Gaussian mixture model-based weighted KICA [6]. As for the representative studies of fault diagnosis, Azim and Sarath [10] extracted independent features using ICA, and then employed long short-term memory (LSTM) networks and artificial neural networks for fine classification, which is suitable for machinery fault diagnosis. Kong et al. [11] extended linear fault reconstruction to kernel dynamic ICA and developed a residual reduced fault subspace (RRFS) extraction method. However, there are still two issues requiring further research.

(1) The fault subspaces extracted from nonlinear models often exhibit excessively high dimensions, leading to significant spatial consumption and a consequent increase in computational complexity.

(2) In practice scenarios, industrial systems often employ redundant designs for critical components. While redundancy improves fault tolerance, it also introduces fault propagation paths and coupling effects between interconnected subsystems. For example, subsystems such as powertrain, braking system, and electronic control system are functionally interdependent. A fault in one subsystem may propagate to others through energy transfer, signal interference, or cascaded control logic, leading to cross-coupled faults. Consequently, the monitored variables (e.g., temperature, pressure, vibration signals) associated with different fault modes often exhibit significant overlaps [12]. These overlaps subsequently lead to commonalities among fault subspaces, introducing uncertainty in the diagnosis process.

As pointed out by Melani et al. [13], no single fault diagnosis method can have all the desired performance, including fast diagnosis capability, robustness, and multiple fault identifiability. The future direction is hybrid solutions, where two or more methods are integrated to complement each other and overcome the limitations of a single method. Motivated by this direction, we attempt to find another diagnosis solution and then integrate it into the framework of the fault reconstruction. Pattern classification, a common type of fault diagnosis method, is a suitable choice. These pattern classification-baseds approaches use historical measurements with different labels to train the model, and then various classifiers, such as Bayesian inference [14, 15], minimax probability machine [16], AdaBoost [17], discriminant function [18], and deep learning [19,20], are employed for fault diagnosis. However, a common problem faced by pattern recognition-based methods in fault diagnosis is the inability to identify unseen faults. Here, the "unseen faults" refers to the faults that are unprecedented in the system and, as a result, has not been included in the model's training process. Correspondingly, faults that have occurred and are included in the training are referred to as "historical faults". Fault reconstruction and pattern classification form a complementary relationship in fault diagnosis. Their respective advantages and disadvantages and application scenarios are summarized in Table 1.

To address the two issues abovementioned, in this study, a reduced reconstruction modeling approach with dual attributes based on KICA (RD-KICA) is proposed for fault diagnosis of nonlinear and non-Gaussian processes. The main innovations of this study are outlined as follows.

(1) *Low-complexity reconstruction modeling with undersampling preprocessing.* Unlike existing KICA variants that directly build models on high-dimensional training data, the RD-KICA introduces an undersampling preprocessing based on the feature vector selection (FVS), and develops a reduced reconstruction modeling, which enables efficient online diagnosis while maintaining detection accuracy.

(2) *Fault representation with dual attributes.* Departing from existing KICA variants that solely rely on fault subspaces for characterization, RD-KICA innovatively represents each fault type with dual attributes: fault subspace and fault magnitude. Specifically, the fault subspace is used for fault reconstruction, while the fault magnitude is used to train a Bayesian classifier. This representation allows the classifiers to re-diagnose in situations where faults cannot be distinguished by fault subspaces.

(3) *A diagnostic strategy fusing reconstruction and pattern classification.* Different from existing KICA methods, which rely solely on fault reconstruction for online diagnosis, RD-KICA designs a unified diagnosis strategy that

fuses fault reconstruction and pattern classification. In this strategy, the fault subspace is first applied for initial diagnosis, and the classifier is activated for re-diagnosis in the presence of uncertainty. This hierarchical diagnosis strategy combines the strengths of both approaches, improving fault diagnosis accuracy, enhancing the ability to deal with unseen faults, and overcoming the uncertainty of traditional KICA methods during the online diagnosis.

The remainder of this paper is structured as follows. Section 2 provides an overview of fault detection based on the KICA. Section 3 details the undersampling and fault library building of the proposed RD-KICA methodology, along with the diagnosis strategy and computational complexity analysis. In Section 4, the simulations and applications conducted on a numerical example, the Tennessee Eastman process (TEP) industrial benchmark, and a rocket servo system are presented. Finally, Section 5 summarizes the conclusion and prospects.

2 Preliminaries

KICA is an extension of ICA that accounts for nonlinearity. A training sample matrix with N samples and m variables is

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N] \in \mathbb{R}^{m \times N}. \tag{1}$$

The core of KICA is mapping the sample matrix $\boldsymbol{X}$ into the high-dimensional feature space $\boldsymbol{\Phi} = [\phi(\boldsymbol{x}_1), \phi(\boldsymbol{x}_2), ..., \phi(\boldsymbol{x}_N)]$, where $\phi(\cdot)$ is a nonlinear mapping function. Due to the difficulty of explicitly expressing $\phi(\cdot)$, the radial basis kernel function is generally used to compute the Gram matrix $\boldsymbol{K} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$, the (i, j)th element of which can be computed as $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/\sigma)$, where σ is the kernel parameter.

After mean centering and scaling, for decorrelation purposes, the Gram matrix $\boldsymbol{K}$ is eigenvalue decomposed as

$$\mu_i \boldsymbol{v}_i = \boldsymbol{K}\boldsymbol{v}_i, \tag{2}$$

where μ_i and $\boldsymbol{v}_i$ are eigenvalue and eigenvector, respectively. Typically, the largest p eigenvalues and their corresponding eigenvectors, which satisfy the condition $\mu_i / \sum_{j=1}^{N} \mu_j > 0.0001$, are used to construct the whitening matrix as

$$\boldsymbol{P} = \sqrt{N}\boldsymbol{\Phi}\boldsymbol{H}\boldsymbol{\Lambda}^{-1}, \tag{3}$$

where $\boldsymbol{H} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_p]$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_p)$. Thus, the whitened score matrix can be defined as

$$\boldsymbol{Z} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{\Phi} = \sqrt{N}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} = \sqrt{N}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{K}. \tag{4}$$

To maximize the non-Gaussianity of independent components, FastICA [21] is implemented to calculate the direction matrix $\boldsymbol{W}$, and the specific procedure can be found in [8]. The ICs of a certain sample $\boldsymbol{x}$ can be expressed as

$$\boldsymbol{s} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{z} = \boldsymbol{W}^{\mathrm{T}}\sqrt{N}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^{\mathrm{T}}[k(\boldsymbol{x}_1, \boldsymbol{x}), ..., k(\boldsymbol{x}_N, \boldsymbol{x})]^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{k}_{\boldsymbol{x}}, \tag{5}$$

where $\boldsymbol{Q} = \sqrt{N}\boldsymbol{W}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{H}^{\mathrm{T}}$ is the demixing matrix, $\boldsymbol{z}$ and $\boldsymbol{k}_{\boldsymbol{x}}$ is whiten score vector and kernel vector, respectively.

To monitor the systematic and nonsystematic changes in nonlinear processes, two statistics are defined as follows:

$$\begin{cases} I^2(\boldsymbol{x}) = \boldsymbol{k}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{k}_{\boldsymbol{x}} = \left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{k}_{\boldsymbol{x}}\right\|^2, \\ SPE(\boldsymbol{x}) = \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})\boldsymbol{z}, \end{cases} \tag{6}$$

where $\boldsymbol{\Sigma} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$ is a coefficient matrix, and $\boldsymbol{I}$ is a unit matrix. Kernel density estimation (KDE) is utilized to determine the control limits I^2_{limit} and SPE_{limit} due to not being limited to any particular distribution.

3 Methodology

In this section, the motivations and implementations of undersampling and fault representation with dual attributes are detailed, along with the fault diagnosis strategy and computational complexity analysis.

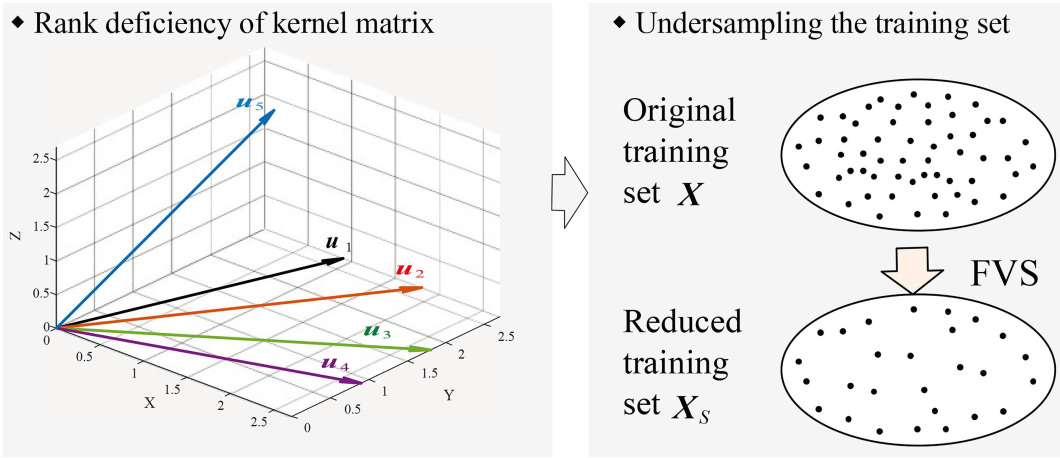

Figure 1 (Color online) Schematic diagram of the FVS.

3.1 Problem statements and motivations

In response to the two issues mentioned in Section 1, it is necessary to investigate a fault diagnosis method that can both reduce computational complexity and improve accuracy, and we may need to delve into the following two questions.

(1) What is the primary factor contributing to the high complexity? How can it be effectively alleviated?

The nonlinear fault subspaces are extracted from the kernel matrix by SVD in the Hilbert space. This manner of extraction makes the fault subspace computationally as large as the number of all training samples even at the testing phase [22], leading to high complexity. Knowing that the kernel matrix is spanned by the raw data, it is reasonable to believe that the kernel matrix is rank-deficient. Thereby, does there exist a basis that approximates all samples in the Hilbert space?

Motivated by the consensus that the generalization capacity relies on the geometric properties of the training data rather than their dimensionality [23], we seek feature vectors (FVs) that can adequately represent all other data. As depicted in Figure 1, in the two-dimensional X-Y plane, any two non-collinear pairs of vectors (e.g., $\boldsymbol{u}_1$ and $\boldsymbol{u}_4$) are sufficient to form a basis representing all other vectors. When the space expands from two-dimensional to three-dimensional X-Y-Z space, it is necessary to combine $\boldsymbol{u}_1$, $\boldsymbol{u}_4$, and a new vector $\boldsymbol{u}_5$ to form a new basis. In this manner, by utilizing these chosen FVs to span a reduced kernel matrix, we can extract subspaces within it, significantly reducing the dimensionality of the subspaces.

(2) When the fault subspace is insufficient for accurate fault diagnosis, are there any other fault attributes that can aid in fault identification?

For different fault types with serious overlap, such as IDV(4) and IDV(11) in TEP, their associated faulty variables also remain consistent [12]. This serious overlap leads to significant commonalities among different subspaces. As a result, the fault effect of one sample can be eliminated by multiple subspaces, as depicted in Figure 2. These samples are known under ill conditions, and their statistics effectively indicate the presence of the fault, as marked by the black line. From Figure 2, it can be found that fault samples can be recovered by both subspaces A and B. This situation leads to uncertainty as to whether these samples belong to fault A or B. In this situation, are there any other fault attributes, apart from the subspace, that can aid in fault identification? Inspired by the fault geometric depiction in [24], i.e., $\boldsymbol{x} = x^* + \boldsymbol{\Xi} f$, where x^* and $\boldsymbol{\Xi} f$ refer to fault-free and faulty-parts, respectively, we can find that there are two key factors responsible for the fault: $\boldsymbol{\Xi}$ and $\boldsymbol{f}$. From the perspective of physical properties, the fault subspace is a low-dimensional feature extracted from fault samples, consisting of multiple singular vectors. Therefore, it indicates the direction information of the fault. The fault magnitude indicates the strength or degree, providing quantitative information to help distinguish the different severity levels of the fault. The combination of the two is complementary, retaining both directional information and increasing magnitude quantification information, thereby improving the accuracy of classification or detection.

As we know, the magnitude $\boldsymbol{f}$ has never received attention in existing research. In this study, we supplement magnitude $\boldsymbol{f}$ as another attribute to assist fault diagnosis, and propose a novel fault representation method with dual attributes.

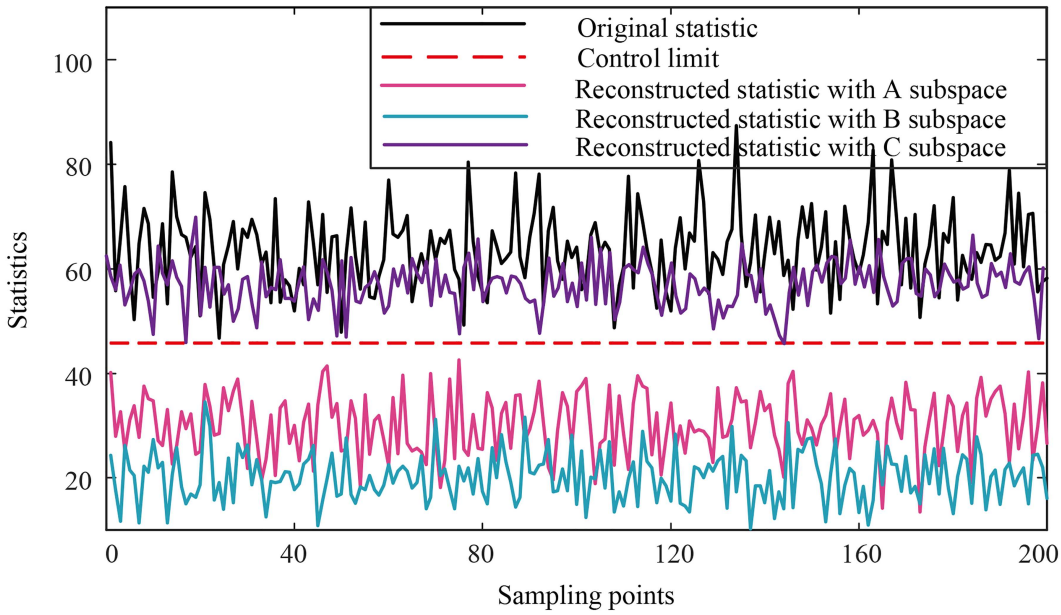

Figure 2 (Color online) The uncertainty situation in fault reconstruction.

3.2 Undersampling the training set using FVS

FVS [25] iteratively selects the samples that are farthest from the span of the current FVs as the next FV, until all samples can be represented as linear combinations of the selected FVs. Such that the information contained in the original training set $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_N]$ is maximally preserved. By doing so, the representative samples, i.e., FVs, can be screened and form a reduced training set $\boldsymbol{X}_S = [\boldsymbol{x}_1^S, ..., \boldsymbol{x}_{N_S}^S]$ ($N_S \leqslant N$). For given FVs $\boldsymbol{X}_S = [\boldsymbol{x}_1^S, ..., \boldsymbol{x}_{N_S}^S]$, the kernel vector of any other samples $\boldsymbol{x}_i$ can be estimated as a linear combination:

$$\hat{\boldsymbol{\varphi}}(\boldsymbol{x}_i) = \boldsymbol{\Phi}_S \cdot \boldsymbol{a}_i, \tag{7}$$

where $\boldsymbol{\Phi}_S = [\phi(\boldsymbol{x}_1^S), ..., \phi(\boldsymbol{x}_{N_s}^S)]$ is the kernel matrix of the $\boldsymbol{X}_S$, and $\boldsymbol{a}_i = [a_{i1}, ..., a_{iNs}]^{\mathrm{T}}$ is the coefficient vector. Next, a factor δ_i [21] is defined to measure collinearity between real kernel vector $\varphi(\boldsymbol{x}_i)$ and estimated kernel vector $\hat{\varphi}(\boldsymbol{x}_i)$:

$$\delta_i = \frac{\|\varphi(\boldsymbol{x}_i) - \hat{\varphi}(\boldsymbol{x}_i)\|^2}{\|\varphi(\boldsymbol{x}_i)\|^2} = \frac{(\varphi(\boldsymbol{x}_i) - \boldsymbol{\Phi}_S \cdot \boldsymbol{a}_i)^{\mathrm{T}}(\varphi(\boldsymbol{x}_i) - \boldsymbol{\Phi}_S \cdot \boldsymbol{a}_i)}{\|\varphi(\boldsymbol{x}_i)\|^2}. \tag{8}$$

Notably, a smaller value of δ_i indicates a smaller angle between $\varphi(\boldsymbol{x}_i)$ and $\hat{\varphi}(\boldsymbol{x}_i)$, implying a stronger collinearity. Hence, the goal is to find the coefficient vector $\boldsymbol{a}_i$ that minimizes the δ_i. Setting the derivation of δ_i to zero yields:

$$\frac{\partial \delta_i}{\partial \boldsymbol{a}_i} = \frac{2(\boldsymbol{\Phi}_S^{\mathrm{T}}\boldsymbol{\Phi}_S)\boldsymbol{a}_i - 2\boldsymbol{\Phi}_S^{\mathrm{T}}\varphi(\boldsymbol{x}_i)}{\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_i)} = 0, \tag{9}$$

$$\frac{\partial \delta_i}{\partial \boldsymbol{a}_i} = 0 \;\Rightarrow\; \boldsymbol{a}_i = (\boldsymbol{\Phi}_S^{\mathrm{T}}\boldsymbol{\Phi}_S)^{-1}\boldsymbol{\Phi}_S^{\mathrm{T}}\varphi(\boldsymbol{x}_i), \tag{10}$$

where $(\boldsymbol{\Phi}_s^{\mathrm{T}}\boldsymbol{\Phi}_S)^{-1}$ exists if the FVs are linearly independent. Substituting (10) into (8), it holds that

$$\min \delta_i = 1 - \frac{\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\Phi}_S(\boldsymbol{\Phi}_S^{\mathrm{T}}\boldsymbol{\Phi}_S)^{-1}\boldsymbol{\Phi}_S^{\mathrm{T}}\varphi(\boldsymbol{x}_i)}{\|\varphi(\boldsymbol{x}_i)\|^2}. \tag{11}$$

Let

$$\boldsymbol{K}_{S,S} = \boldsymbol{\Phi}_S^{\mathrm{T}}\boldsymbol{\Phi}_S = \{k(\boldsymbol{x}_p^S, \boldsymbol{x}_q^S)\}, \tag{12}$$

$$\boldsymbol{k}_{S,\boldsymbol{x}_i} = \boldsymbol{\Phi}_S^{\mathrm{T}}\varphi(\boldsymbol{x}_i) = \{k(\boldsymbol{x}_p^S, \boldsymbol{x}_i)\}, \tag{13}$$

where $1 \leqslant p \leqslant N_S, 1 \leqslant q \leqslant N_S$, $\boldsymbol{K}_{S,S}$ refers to the matrix of dot products of FVs, and $\boldsymbol{k}_{S,\boldsymbol{x}_i}$ refers to the vector of dot products between $\boldsymbol{x}_i$ and FVs. Thus, Eq. (11) can be rewritten as

$$\min \delta_i = 1 - \frac{\boldsymbol{k}_{S,\boldsymbol{x}_i}^{\mathrm{T}} \boldsymbol{K}_{S,S}^{-1} \boldsymbol{k}_{S,\boldsymbol{x}_i}}{k_{\boldsymbol{x}_i,\boldsymbol{x}_i}}. \tag{14}$$

The goal is to find the set $\boldsymbol{X}_S$ that minimizes (14) over all samples $\boldsymbol{x}_i \in X$:

$$\min_{\boldsymbol{X}_S} \left(\sum_{\boldsymbol{x}_i \in \boldsymbol{X}} \left(1 - \frac{\boldsymbol{k}_{S,\boldsymbol{x}_i}^{\mathrm{T}} \boldsymbol{K}_{S,S}^{-1} \boldsymbol{k}_{S,\boldsymbol{x}_i}}{k_{\boldsymbol{x}_i,\boldsymbol{x}_i}} \right) / N \right). \tag{15}$$

Define the global fitness J_S^* and local fitness $J_S(\boldsymbol{x}_i)$ for a given $\boldsymbol{X}_S$ as

$$J_S^* = \frac{1}{N} \sum_{\boldsymbol{x}_i \in \boldsymbol{X}} J_S(\boldsymbol{x}_i), \tag{16}$$

$$J_S(\boldsymbol{x}_i) = \frac{\boldsymbol{k}_{S,\boldsymbol{x}_i}^{\mathrm{T}} \boldsymbol{K}_{S,S}^{-1} \boldsymbol{k}_{S,\boldsymbol{x}_i}}{k_{\boldsymbol{x}_i,\boldsymbol{x}_i}}. \tag{17}$$

Then, the goal of (15) can be transformed as

$$\max_{\boldsymbol{X}_S} \left(J_S^* \right). \tag{18}$$

The selection of the set $\boldsymbol{X}_S$ is iterative and follows a sequential forward selection, which tends to select FVs that are close to orthogonality. The pseudo-code of this procedure is provided in Algorithm 1. In the first iteration, we look for the sample that yields the maximum J_S^*. For all subsequent iterations, the next FVs are determined by selecting the sample that provides the minimum $J_S(\boldsymbol{x}_i)$. The iterative selection procedure terminates when $\boldsymbol{K}_{S,S}$ is not invertible, indicating that the current $\boldsymbol{X}_S$ is a good approximation basis for the original training set $\boldsymbol{X}$ [25].

Since FVS significantly reduces the data size, it is used for undersampling the training data without losing the informative data, therefore reducing the computational complexity.

Algorithm 1 Iterative selection procedure of the FVS.

Input: original training set $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_N]$, reduced training set $\boldsymbol{X}_S = [\,]$, temporary vector $\boldsymbol{J}_{S\_all} = [\,]$.
Steps:
First FV:
1. **for** $i = 1$ to N;
2. $\boldsymbol{X}_S = \{\boldsymbol{x}_i\}$, calculate J_S^* with respect to the present $\boldsymbol{X}_S$, and restore it into a vector: $\boldsymbol{J}_{S\_all}(i) \leftarrow J_S^*$;
3. **endfor**
4. Select the sample $\boldsymbol{x}_k$ that yields the maximum J_S^* as the first FV, and add it into $\boldsymbol{X}_S$: $\boldsymbol{X}_S = [\boldsymbol{x}_k : J_S^*(\boldsymbol{x}_k) = \max(\{\boldsymbol{J}_{S\_all}(j)\})]$;
5. Delete the FVs from $\boldsymbol{X}$: $\boldsymbol{X} = \boldsymbol{X} \backslash \boldsymbol{X}_S$;
Subsequent FVs:
6. **while** $\boldsymbol{K}_{S,S}$ is invertible;
7. Calculate the $J_S(\boldsymbol{x}_i)$ for all samples $\boldsymbol{x}_i$ in $\boldsymbol{X}$ with respect to the present $\boldsymbol{X}_S$;
8. Select the sample $\boldsymbol{x}_k$ that provides the minimum $J_S(\boldsymbol{x}_i)$: $J_S(\boldsymbol{x}_k) = \min(\{J_S(\boldsymbol{x}_i)\}), \boldsymbol{x}_i \in \boldsymbol{X}$;
9. Add $\boldsymbol{x}_k$ into set $\boldsymbol{X}_S$: $\boldsymbol{X}_S = \{\boldsymbol{X}_S, \boldsymbol{x}_k\}$;
10. Delete FVs from $\boldsymbol{X}$: $\boldsymbol{X} = \boldsymbol{X} \backslash \boldsymbol{X}_S$;
11.**endwhile**.
Output: reduced training set $\boldsymbol{X}_S$.

3.3 Fault library building with dual attributes

When different fault types encounter serious overlap, there exists uncertainty in identifying the specific fault type. To tackle this issue, a fault library with dual attributes is built in this subsection. In contrast to traditional reconstruction-based methods, we not only extract the fault subspace for each fault type, but also the fault magnitudes, which are utilized to train a classifier. In this study, we choose Bayesian binary classifier (BBC) due to its simple implementation and high interpretability. It also allows us to dynamically adjust candidate fault types based on the reconstructed results during online diagnosis. This adaptability makes it particularly suitable for the fault re-diagnosis task in our study. This approach endows each fault type with dual diagnostic bases: fault subspace and BBC. When it is difficult to make an accurate diagnosis only relying on fault subspace, BBC can be used for re-diagnosis, so as to enhance the accuracy and reliability of fault diagnosis.

3.3.1 *Extraction of fault subspaces*

Supposing the kth type fault set is $\boldsymbol{X}_f^{(k)}(k=1,...,c)$, and its kernel matrix is $\boldsymbol{K}_f^{(k)}$, then the independent subspace can be expressed as

$$\hat{\boldsymbol{K}}_f^{(k)} = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{K}_f^{(k)}, \tag{19}$$

where $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{W}/\sqrt{N}$ is the mixing matrix. Perform SVD on $\hat{\boldsymbol{K}}_f^{(k)}$ as follows:

$$\hat{\boldsymbol{K}}_f^{(k)} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{S}, \tag{20}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, ..., \boldsymbol{u}_{N_S}]$ is the left singular matrix. Notably, $\boldsymbol{u}_i$ is arranged in descending order according to singular values. $\boldsymbol{D}$ and $\boldsymbol{S}$ are the singular value matrix and the right singular matrix, respectively. Fault subspace consists of the $\boldsymbol{u}_i$ corresponding to the first θ_k singular vectors, i.e., $\boldsymbol{\Xi}_k = [\boldsymbol{u}_1, ..., \boldsymbol{u}_{\theta_k}]$. For the fault sample $\boldsymbol{x}$, its kernel vector can be formulated as a sum of fault-free part $\boldsymbol{k}^*$ and faulty part $\boldsymbol{\Xi}_k\boldsymbol{f}$:

$$\boldsymbol{k} = \boldsymbol{k}^* + \boldsymbol{\Xi}_k\boldsymbol{f}, \tag{21}$$

where $\boldsymbol{\Xi}_k \in \mathbb{R}^{N_S\times\theta_k}$ and $\boldsymbol{f} \in \mathbb{R}^{\theta_k\times 1}$ are the fault subspace and magnitude, respectively. Due to the assumption that $\boldsymbol{k}^*$ is the fault-free part, its statistic should be as small as possible. In this study, we take the I^2 statistic as an example, and the estimation of $\boldsymbol{f}$ can be regarded as an extremum problem:

$$J(\boldsymbol{f}) = \arg\min_{\boldsymbol{f}}\{I^2(\boldsymbol{k}^*)\} = \arg\min_{\boldsymbol{f}}\left\{\left\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{k}-\boldsymbol{\Xi}_k\boldsymbol{f})\right\|^2\right\}, \tag{22}$$

which is an unconstrained least-squares issue with the following analytical solution [3]:

$$\boldsymbol{f} = (\boldsymbol{\Xi}_k^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Xi}_k)^{\dagger}\boldsymbol{\Xi}_k^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{k}. \tag{23}$$

θ_k should be set to the minimum dimension that makes the fault reconstruction rate (FRR) up to a default ς, and the FRR is defined as

$$\mathrm{FRR} = N_{fn}/N_f \ \times 100\%, \tag{24}$$

where N_{fn} represents the number of fault samples recovered, and N_f refers to the total number of fault samples.

3.3.2 *Construction of Bayesian binary classifiers*

After extracting the fault subspace, this subsection details how to construct a BBC based on fault magnitudes, taking the kth fault as an example. First, we reconstruct all historical faulty samples with the kth fault subspace $\boldsymbol{\Xi}_k$ according to (23), and record their corresponding magnitudes into two vectors. The fault magnitudes for the kth and non-kth fault samples are defined as follows:

$$\boldsymbol{F}_k = [\boldsymbol{f}_{k,1}, ..., \boldsymbol{f}_{k,N_k}] \in \mathbb{R}^{\theta_k\times N_k}, \tag{25}$$

$$\boldsymbol{F}_{nk} = [\boldsymbol{f}_{nk,1}, ..., \boldsymbol{f}_{nk,N_{nk}}] \in \mathbb{R}^{\theta_k\times N_{nk}}, \tag{26}$$

where $\boldsymbol{f}_{k,i}$ denotes the fault magnitude corresponding to the $\boldsymbol{\Xi}_k$ of the ith sample labeled as the kth fault, and $\boldsymbol{f}_{nk,i}$ denotes that of the ith sample labeled as non-kth fault; N_k and N_{nk} represent the number of historical labeled samples of the kth and non-kth fault, respectively. Second, to ensure the balance of the positive and negative samples, we randomly select $\min\{N_k, N_{nk}\}$ samples from $\boldsymbol{F}_k/\boldsymbol{F}_{nk}$ as positive/negative samples, which are still defined as $\boldsymbol{F}_k$ and $\boldsymbol{F}_{nk}$ for ease of presentation. Such a treatment can avoid the degradation of classifier performance caused by imbalance. Finally, $\boldsymbol{F}_k$ and $\boldsymbol{F}_{nk}$ are used as positive and negative samples to construct the BBC, which can provide a probability of belonging to the kth fault for each sample.

Assuming that the fault magnitude of sample $\boldsymbol{x}$ reconstructed by subspace $\boldsymbol{\Xi}_k$ is $\boldsymbol{f}$, the probability of which belongs to the kth fault can be expressed as

$$p(\boldsymbol{x}\in c_k) = \frac{p(\boldsymbol{f}|c_k)p(c_k)}{p(\boldsymbol{f}|c_k)p(c_k)+p(\boldsymbol{f}|c_{nk})p(c_{nk})}, \tag{27}$$

where $p(c_k)$ and $p(c_{nk})$ denote the prior probabilities of positive and negative samples, respectively, which are equal to 0.5 in this study since samples have been balanced as described in the previous paragraph. The $p(\boldsymbol{f}|c_k)$ and

$p(\boldsymbol{f}|c_{nk})$ denote the conditional probability density of positive and negative samples, respectively. Since the specific distributions of $\boldsymbol{F}_k$ and $\boldsymbol{F}_{nk}$ are unknown, we employ the Parzen window [26], a nonparametric KDE method, to estimate the probability density as follows:

$$p(\boldsymbol{f}|c_k) = \frac{\sum_{i=1}^{N_k^*} \exp\left(-\frac{1}{2}(\boldsymbol{f}-\boldsymbol{f}_{k,i})^{\mathrm{T}}\boldsymbol{S}_{F_k}^{-1}(\boldsymbol{f}-\boldsymbol{f}_{k,i})\right)}{N_k^* \times \sqrt{(2\pi)^{\theta_k} \times \left|\boldsymbol{S}_{\boldsymbol{F}_k}\right|}}, \tag{28}$$

$$p(\boldsymbol{f}|c_{nk}) = \frac{\sum_{i=1}^{N_k^*} \exp\left(-\frac{1}{2}(\boldsymbol{f}-\boldsymbol{f}_{nk,i})^{\mathrm{T}}\boldsymbol{S}_{F_{nk}}^{-1}(\boldsymbol{f}-\boldsymbol{f}_{nk,i})\right)}{N_k^* \times \sqrt{(2\pi)^{\theta_k} \times \left|\boldsymbol{S}_{F_{nk}}\right|}}, \tag{29}$$

where $N_k^* = \min\{N_k, N_{nk}\}$, $\boldsymbol{S}_{F_k}$ and $\boldsymbol{S}_{F_{nk}}$ are the covariance matrices of $\boldsymbol{F}_k$ and $\boldsymbol{F}_{nk}$, respectively. Thus, we can construct a BBC for the kth fault, defined as Π_k.

Up to this point, each type of fault has a fault subspace and a BBC associated with it, and a complete fault library, denoted as $\{\{\boldsymbol{\Xi}_1, \Pi_1\}, ..., \{\boldsymbol{\Xi}_c, \Pi_c\}\}$, can be built. This enhanced fault library will further guide the subsequent online fault diagnosis.

3.4 Online fault diagnosis strategy

For a given fault sample $\boldsymbol{x}_{\mathrm{new}}$, its centered and scaled kernel vector is defined as $\boldsymbol{k}_{\mathrm{new}}$. For $\boldsymbol{\Xi}_k$, according to (23) and (6), its fault magnitude and the reconstructed statistic can be calculated as

$$\boldsymbol{f}_{\mathrm{new},k} = (\boldsymbol{\Xi}_k^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Xi}_k)^{\dagger}\boldsymbol{\Xi}_k^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{k}_{\mathrm{new}}, \tag{30}$$

$$I_{(k)}^2(\boldsymbol{k}_{\mathrm{new}}^*) = \left\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{k}_{\mathrm{new}} - \boldsymbol{\Xi}_k\boldsymbol{f}_{\mathrm{new},k})\right\|^2. \tag{31}$$

Following (30) and (31), we reconstruct $\boldsymbol{k}_{\mathrm{new}}$ in turn with $\{\boldsymbol{\Xi}_1, ..., \boldsymbol{\Xi}_c\}$, and record all reconstructed I^2 statistic as

$$\boldsymbol{I}_{\mathrm{new}} = \left[I_{(1)}^2(\boldsymbol{k}_{\mathrm{new}}^*), ..., I_{(c)}^2(\boldsymbol{k}_{\mathrm{new}}^*)\right]. \tag{32}$$

To determine the fault type, it is necessary to count the number of subspaces in which the sample $\boldsymbol{x}_{\mathrm{new}}$ can be reconstructed below the I_{limit}^2:

$$C_{\mathrm{new}} = \sum_{k=1}^{c} \mathbb{I}\left(I_{(k)}^2(\boldsymbol{k}_{\mathrm{new}}^*) < I_{\mathrm{limit}}^2\right), \tag{33}$$

where $\mathbb{I}(\cdot)$ is an indicator function. If it holds that $I_{(k)}^2(\boldsymbol{k}_{\mathrm{new}}^*) < I_{\mathrm{limit}}^2$, then $\mathbb{I}\left(I_{(k)}^2(\boldsymbol{k}_{\mathrm{new}}^*) < I_{\mathrm{limit}}^2\right) = 1$, otherwise, $\mathbb{I}\left(I_{(k)}^2(\boldsymbol{k}_{\mathrm{new}}^*) < I_{\mathrm{limit}}^2\right) = 0$.

There are following three cases when analyzing C_{new}.

(1) $C_{\mathrm{new}} = 0$. This means that these subspaces cannot effectively reflect the fault information of $\boldsymbol{x}_{\mathrm{new}}$. In this case, we can consider the sample $\boldsymbol{x}_{\mathrm{new}}$ as an unseen fault. In order to ensure the completeness of the fault library, the system should trigger an "unseen fault" alert. It is then necessary to remind the practitioner to collect fault samples and extract the corresponding fault subspaces and fault magnitudes, which are then integrated into the fault library, enabling dynamic updates and progressive improvement.

(2) $C_{\mathrm{new}} = 1$. This means that only one subspace can reconstruct $\boldsymbol{x}_{\mathrm{new}}$ below the control limit I_{limit}^2, which indicates that this subspace can uniquely reflect the fault characteristics. Therefore, we can classify $\boldsymbol{x}_{\mathrm{new}}$ as the fault type corresponding to this subspace, i.e.,

$$\Gamma(\boldsymbol{x}_{\mathrm{new}}) = \left\{k \,\middle|\, \mathbb{I}\left(I_{(k)}^2(\boldsymbol{k}_{\mathrm{new}}^*) < I_{\mathrm{limit}}^2\right) = 1\right\}, \tag{34}$$

where $\Gamma(\boldsymbol{x}_{\mathrm{new}})$ refers to the fault type of $\boldsymbol{x}_{\mathrm{new}}$.

(3) $C_{\mathrm{new}} > 1$. This means that multiple subspaces can reconstruct $\boldsymbol{x}_{\mathrm{new}}$ below the control limit I_{limit}^2. In this case, further analysis based on the BBC in the fault library is needed. Assume that $\boldsymbol{x}_{\mathrm{new}}$ can be reconstructed below the I_{limit}^2 by $\left\{\boldsymbol{\Xi}^{(1)}, ..., \boldsymbol{\Xi}^{(\alpha)}\right\}$, and the corresponding magnitudes are recorded as $\boldsymbol{f}_{\mathrm{new}}^{(1)}, ..., \boldsymbol{f}_{\mathrm{new}}^{(\alpha)}$. Notably, $\boldsymbol{\Xi}^{(1)}$ here is not the same as $\boldsymbol{\Xi}_1$. $\boldsymbol{\Xi}^{(1)}$ represents the first subspace that can reconstruct $\boldsymbol{x}_{\mathrm{new}}$ below the I_{limit}^2, and it can

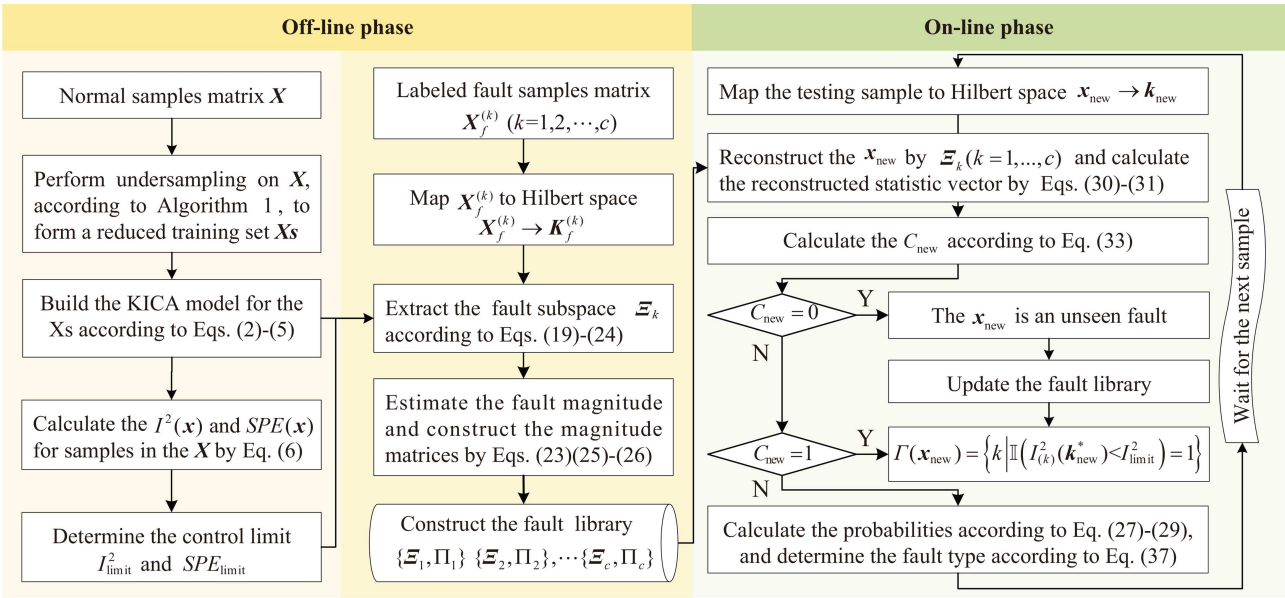

Figure 3 (Color online) Flowchart of the proposed RD-KICA.

Table 2 The number of flops needed during online fault diagnosis.

Main time-consuming steps	KICA	RD-KICA
Calculate kernel vector $\boldsymbol{k}_{\text{new}}$ and scale it	$O(mN+N^2)$	$O(mN_S+N_S^2)$
Calculate the I^2 for the $\boldsymbol{k}_{\text{new}}$	$O(N^2)$	$O(N_S^2)$
Calculate the fault magnitude $\boldsymbol{f}_{\text{new},k}$	$O(N^2+N^2\theta_k+N\theta_k^2+\theta_k^3)$	$O(N_S^2+N_S^2\theta_k+N_S\theta_k^2+\theta_k^3)$
Reconstruct the $\boldsymbol{k}_{\text{new}}$ and calculate the reconstructed I^2 statistic	$O(N^2+N\theta_k)$	$O(N_S^2+N_S\theta_k)$
Identify the fault using BBCs	–	$O(N_k\theta_k^2)$

be any of $\{\boldsymbol{\Xi}_1,...,\boldsymbol{\Xi}_c\}$. Subsequently, based on the trained BBCs, the probability of $\boldsymbol{x}_{\text{new}}$ belonging to each fault type is calculated as follows.

(i) For each candidate fault type, magnitudes $\boldsymbol{f}_{\text{new}}^{(1)},...,\boldsymbol{f}_{\text{new}}^{(\alpha)}$ are substituted into (27)–(29) to calculate the posterior probabilities $\boldsymbol{p}_{\text{new}}=\left[p(\boldsymbol{x}_{\text{new}}\in c^{(1)}),...,p(\boldsymbol{x}_{\text{new}}\in c^{(\alpha)})\right]$. For example, if $\boldsymbol{x}_{\text{new}}$ can be reconstructed below I^2_{limit} by $\boldsymbol{\Xi}_2$ and $\boldsymbol{\Xi}_4$, then the candidate fault types are fault 2 and 4, and it holds that $\boldsymbol{\Xi}^{(1)}=\boldsymbol{\Xi}_2$, $\boldsymbol{\Xi}^{(2)}=\boldsymbol{\Xi}_4$, $\boldsymbol{f}_{\text{new}}^{(1)}=\boldsymbol{f}_{\text{new},2}$, $\boldsymbol{f}_{\text{new}}^{(2)}=\boldsymbol{f}_{\text{new},4}$, $p(\boldsymbol{x}_{\text{new}}\in c^{(1)})=p(\boldsymbol{x}_{\text{new}}\in c_2)$, and $p(\boldsymbol{x}_{\text{new}}\in c^{(2)})=p(\boldsymbol{x}_{\text{new}}\in c_4)$. According to (27), probabilities of $\boldsymbol{x}_{\text{new}}$ belonging to fault 2 and 4 are computed as

$$p(\boldsymbol{x}_{\text{new}}\in c_2)=\frac{p(\boldsymbol{f}_{\text{new},2}|c_2)p(c_2)}{p(\boldsymbol{f}_{\text{new},2}|c_2)p(c_2)+p(\boldsymbol{f}_{\text{new},2}|c_{n2})p(c_{n2})}, \tag{35}$$

$$p(\boldsymbol{x}_{\text{new}}\in c_4)=\frac{p(\boldsymbol{f}_{\text{new},4}|c_4)p(c_4)}{p(\boldsymbol{f}_{\text{new},4}|c_4)p(c_4)+p(\boldsymbol{f}_{\text{new},4}|c_{n4})p(c_{n4})}. \tag{36}$$

(ii) By comparing the posterior probabilities, we can classify $\boldsymbol{x}_{\text{new}}$ as the fault type corresponding to the maximum posterior probability, i.e.,

$$\Gamma(\boldsymbol{x}_{\text{new}})=\arg\max_{1\leqslant i\leqslant\alpha}\left\{p(\boldsymbol{x}_{\text{new}}\in c^{(i)})\right\}. \tag{37}$$

Overall, the flowchart of the proposed RD-KICA is summarized in Figure 3.

3.5 Computational complexity analysis

Given that the training phase can be executed offline, the computational costs during the online phase using subspaces extracted by RD-KICA and traditional KICA are primarily analyzed. The number of floating-point basic operations (flops) of the main time-consuming steps for these two methods is calculated and recorded in Table 2.

After undersampling with FVS, the kernel matrix is reduced from $N\times N$ to $N_S\times N_S$ dimension, which leads to the subspace dimension extracted from it being also significantly reduced. It is worth noting that there exists $N>N_S>\theta_k$ in Table 2, and '–' means that this step is not required for the corresponding method. Although

RD-KICA requires an additional $O(N_k\theta_k^2)$ flops in the presence of uncertainty during diagnosis, its computational complexity is still reduced compared to KICA.

4 Simulation and application studies

In this section, a nonlinear numerical example, the TEP, and the rocket servo system were used to validate the performance of the proposed RD-KICA method. We compared the proposed method with seven methods: KSFDA [17], ENBC [14], CNN-LSTM, RRFS [11], LPP-Bayesian, PCA-KNN, and RD-KICA-without-FVS. Notably, LPP-Bayesian and PCA-KNN, respectively, use LPP (local preserved projection) and PCA techniques to reduce the dimensionality of samples, extracting key features, and then combine with corresponding classifiers to classify the reduced samples; while the RRFS is a variant of KICA. The RD-KICA-without-FVS refers to the ablation version of the proposed method without FVS undersampling, and is used to perform ablation experiments to test the influence of FVS undersampling on computational efficiency and diagnostic accuracy. Finally, we implemented online diagnosis on a rocket servo system, detailing the diagnostic procedure of the proposed RD-KICA. Notably, the confidence level and the default reconstruction ratio are set to 0.95 in all following experiments.

4.1 Numerical simulation

Revisit the nonlinear and non-Gaussian numerical example [27] and revise it as follows:

$$\begin{aligned}
x_1 &= t_1 + e_1,\\
x_2 &= t_1^2 - 2t_2 + e_2,\\
x_3 &= -t_2^3 + 3t_3^2 + e_3,\\
x_4 &= t_2 + t_4 + e_4,\\
x_5 &= t_3^2 - 2t_4 + e_5,\\
x_6 &= -t_3^2 + 3t_4^3 + e_6,\\
x_7 &= t_1^2 - 3t_3 + e_7,
\end{aligned} \tag{38}$$

where $t_i \sim U(0,1)$ and $e_i \sim N(0, 0.01^2)$. Based on (38), we set the fault as follows:

$$\boldsymbol{x} = \boldsymbol{x}^* + \lambda\boldsymbol{\Theta}, \tag{39}$$

where $\boldsymbol{x}^* = [x_1, x_2, ..., x_7]$ denotes the normal data, and $\boldsymbol{x}$ denotes the faulty data with magnitude λ added to $\boldsymbol{x}^*$ in the $\boldsymbol{\Theta}$ direction. In this simulation, we set the following 5 fault scenarios:

$$\begin{aligned}
&\text{Fault 1: } \boldsymbol{\Theta} = [\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ ]^{\mathrm{T}},\ \lambda_1 = 2,\\
&\text{Fault 2: } \boldsymbol{\Theta} = [\ 2\ 0\ 0\ 3\ 0\ 0\ 0\ ]^{\mathrm{T}},\ \lambda_2 = 1,\\
&\text{Fault 3: } \boldsymbol{\Theta} = [\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ ]^{\mathrm{T}},\ \lambda_3 = 5,\\
&\text{Fault 4: } \boldsymbol{\Theta} = [\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ ]^{\mathrm{T}},\ \lambda_4 = 4,\\
&\text{Fault 5: } \boldsymbol{\Theta} = [\ 0\ 1\ 0\ 1\ 0\ 1\ 2\ ]^{\mathrm{T}},\ \lambda_5 = 4.
\end{aligned} \tag{40}$$

Eq. (40) shows that both Faults 1 and 2 are associated with variable $\boldsymbol{x}_1$, while Faults 3 and 4 share the same fault direction, with the sole difference lying in the fault magnitudes. As for Fault 5, it overlaps partially with other faults in certain fault variables. Such a fault setup aims to assess the diagnostic capability and accuracy of different methods in the scenario of overlapping fault variables.

In this simulation, Faults 1–4 are served as historical faults, while Fault 5 is served as an unseen fault (marked with $*$ in the subsequent tables and graphs figures). We generated 800 normal samples based on (38) for building the model. Meanwhile, according to (39), 800 samples were generated for each historical fault, from which 300 samples were randomly selected for extracting the fault subspace, and the remaining 500 samples were used to test the diagnosis performance. For Fault 5, 500 fault samples were generated to test the feedback ability of different methods toward unseen samples. The detailed fault diagnosis confusion matrices are presented in Figure 4. Note that in the confusion matrix of the RRFS, there is a key concept "Uncertain". This concept refers to the situation where samples can be recovered by multiple subspaces. Consequently, RRFS fails to distinguish the fault type of these samples, resulting in poor diagnosis accuracy in Table 3. As can be seen from Table 3, ENBC, CNN-LSTM,

Table 3 Fault diagnosis accuracy on numerical simulation (%). The best results are in bold.

Fault ID		KSFDA	ENBC	CNN-LSTM	RRFS	LPP-Bayesian	PCA-KNN	RD-KICA-without-FVS	RD-KICA
Historical	1	77.80	92.20	95.20	9.00	97.60	**98.60**	90.40	90.60
	2	99.60	99.00	**100.00**	86.20	97.20	99.20	96.60	96.20
	3	70.00	90.60	89.40	85.40	85.00	82.80	86.20	**92.80**
	4	58.00	83.60	80.60	0.00	81.40	77.40	**91.00**	86.20
	Ave.	76.35	91.35	91.30	45.14	90.30	89.50	91.05	**91.45**
Unseen	5*	0.00	0.00	0.00	87.00	0.00	0.00	92.20	**92.40**

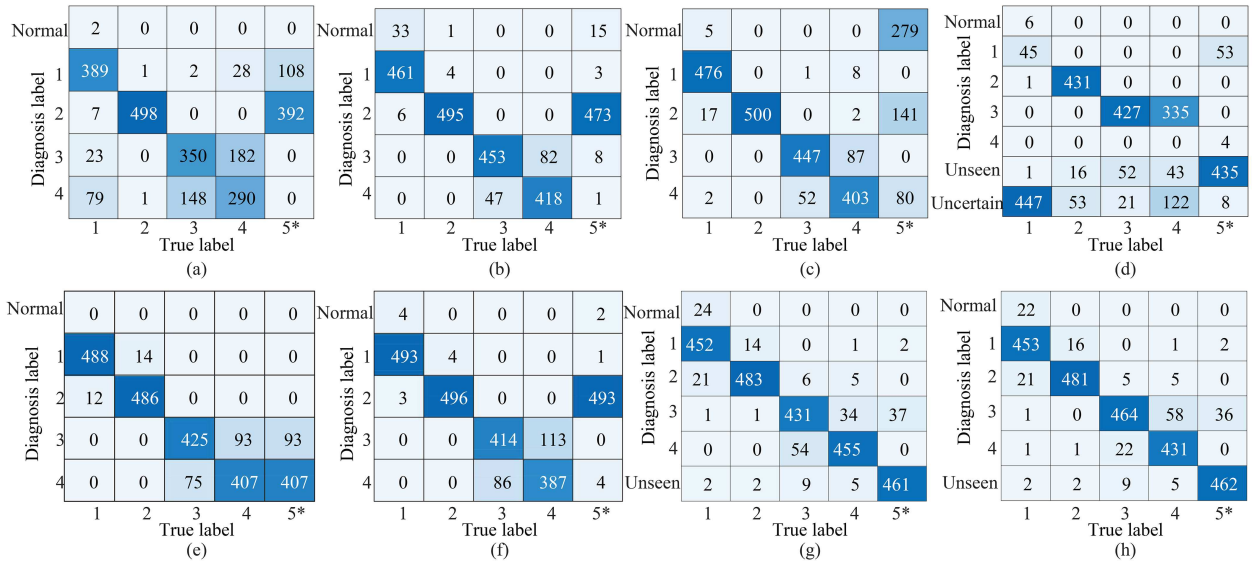

Figure 4 (Color online) Diagnostic confusion matrix of the numerical example. (a) KSFDA; (b) ENBC; (c) CNN-LSTM; (d) RRFS; (e) LPP-Bayesian; (f) PCA-KNN; (g) RD-KICA-without-FVS; (h) RD-KICA.

LPP-Bayesian, and PCA-KNN all show high diagnostic accuracy for historical faults, but all these methods based on pattern recognition fail to identify unseen faults; while RRFS can effectively identify unseen faults, but its average diagnostic accuracy for historical faults is only 45.14%. In contrast, the proposed RD-KICA can identify 92.40% of unseen faults while maintaining high diagnostic accuracy for historical faults.

4.2 Applications on TEP

The TEP [28] is a well-known benchmark process for validating fault detection and diagnosis performance. The TEP encompasses 21 types of faults, denoted as IDV(1)-IDV(21). In this study, IDVs(2, 4, 8, 11, and 14) are regarded as historical faults, while IDV(18) is treated as an unseen fault. Both IDV(2) and IDV(8) are related to the feed concentration in Stream 4, while IDVs(4, 11, and 14) are all associated with the reactor cooling water [2,12]. It is anticipated that samples from these faults may exhibit significant overlap, hence their selection for diagnostic accuracy verification.

A total of 1460 normal samples were collected, with 960 samples allocated for the training. These training samples underwent undersampling to create a reduced training set comprising 195 samples, which was used to establish the KICA model. The remaining 500 samples were used to compute control limits. Furthermore, a fault library with dual attributes was constructed, involving 480 samples for each fault type. The parameters are configured as follows: for KSFDA, the tuning parameter is chosen as $\beta = 0.5$, and the kernel parameter is set to $\sigma = 5.4$; these values are taken from the original literature [17]. Using cross-validation, we set the number of neighbors of LPP-Bayesian as 100. The number of principal components in PCA-KNN is determined by the cumulative variance contribution rate. For RRFS and RD-KICA, the kernel parameter is set to $\sigma = 3052$, which is the result of tabu optimization. The CNN-LSTM is composed of a three-layer convolution network and a two-layer LSTM network. The number of convolution kernels is 32, 64, and 128 with size 1×3, and the number of unit nodes in the LSTM layer is 128.

Table 4 Fault diagnosis accuracy on TEP (%).

Fault ID		KSFDA	ENBC	CNN-LSTM	RRFS	LPP-Bayesian	PCA-KNN	RD-KICA-without-FVS	RD-KICA
Historical	2	95.88	97.25	**98.50**	6.25	98.00	97.63	97.00	97.00
	4	88.50	**100.00**	99.75	0.38	45.00	84.63	97.63	99.25
	8	91.63	64.00	66.00	49.00	**96.75**	74.75	81.63	91.13
	11	53.00	30.50	52.63	**69.25**	42.75	56.25	43.88	50.50
	14	92.88	**99.75**	78.75	99.50	34.63	66.50	96.63	**99.75**
	Ave	84.38	78.30	79.13	44.88	63.43	75.89	83.35	**87.53**
Unseen	18*	0.00	0.00	0.00	77.13	0.00	0.00	**83.75**	82.88

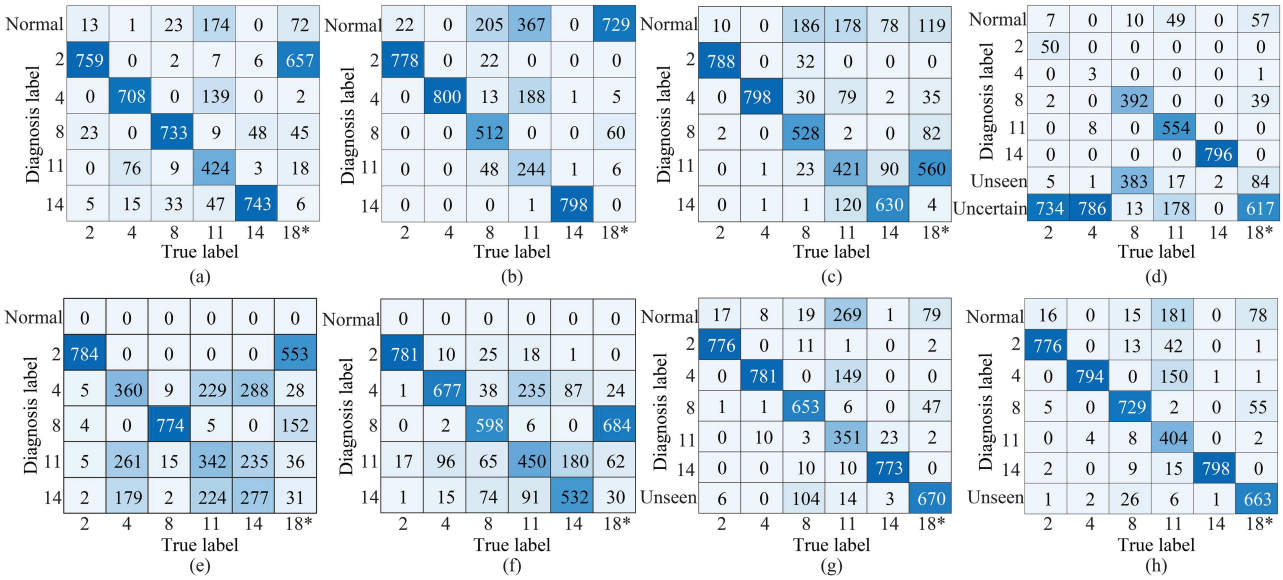

Figure 5 (Color online) Diagnostic confusion matrix of the TEP. (a) KSFDA; (b) ENBC; (c) CNN-LSTM; (d) RRFS; (e) LPP-Bayesian; (f) PCA-KNN; (g) RD-KICA-without-FVS; (h) RD-KICA.

4.2.1 *Experiments on fault diagnosis accuracy*

To clearly show the fault diagnosis results of different methods, we visualize them as confusion matrices in Figure 5, and summarize testing accuracy in Table 4. Diagnostic outcomes of KSFDA, ENBC, CNN-LSTM, LPP-Bayesian, and PCA-KNN show high diagnostic accuracy for historical faults. However, since they lack an identification mechanism for unseen faults, all samples of IDV(18) are misdiagnosed. In contrast, RRFS can effectively identify 77.13% unseen fault samples, while its diagnostic accuracy for historical faults is poor. In contrast, the proposed RD-KICA combines the advantages of reconstruction and classification to achieve the highest average accuracy, as well as the ability to identify unseen faults.

Although the diagnostic accuracy of RD-KICA shown in Table 4 is not the highest among all fault cases, it exhibits the greatest stability. Compared to the RRFS, the average diagnostic accuracy of RD-KICA for historical faults is improved by 42.65%, indicating that adding magnitude as another attribute contributes to enhanced diagnostic accuracy. Compared to KSFDA, ENBC, CNN-LSTM, LPP-Bayesian, and PCA-KNN, RD-KICA can effectively identify 82.88% of unseen faults while ensuring comparable diagnostic accuracy for historical faults. Moreover, the diagnostic accuracy of RD-KICA is comparable to, or even slightly higher than, that of RD-KICA-without-FVS. These findings suggest that undersampling the training set does not compromise fault diagnostic accuracy.

4.2.2 *Experiments on computational complexity*

In this subsection, we compare the computing time (CT) among six methods. CT is the time taken to diagnose each sample online. We executed the online diagnosis programs of the different methods 60 times on the same computer equipped with an Intel Core$^{\text{TM}}$ i7-1165G7 processor (2.8 GHz) and 16 GB of RAM. The results of these methods are presented in Table 5.

Unsurprisingly, ENBC exhibits the highest CT since it classifies high-dimensional samples without dimensionality reduction. Both RD-KICA-without-FVS and RRFS, which utilize nonlinear models for fault reconstruction,

Table 5 Computing time (CT) (ms) of different methods for TEP.

Fault ID	KSFDA	ENBC	CNN-LSTM	RRFS	LPP-Bayesian	PCA-KNN	RD-KICA-without-FVS	RD-KICA
2	0.9262	3.9496	1.0175	1.1556	**0.0847**	0.2750	1.1279	0.4896
4	0.9564	3.7932	1.0284	1.1975	**0.0823**	0.2702	2.3804	1.1625
8	0.9398	3.8355	1.0781	1.4395	**0.0788**	0.2714	1.0984	0.2145
11	0.9405	3.8842	1.0349	1.4317	**0.0757**	0.2611	1.6609	0.8842
14	0.9491	4.4864	1.0037	1.5262	**0.0771**	0.2604	1.1591	0.3461
18*	1.0197	4.5781	1.0330	1.4998	**0.0874**	0.2609	1.1289	0.1416
Ave.	0.9553	4.0878	1.0326	1.3751	**0.0810**	0.2665	1.4259	0.5398

Table 6 Fault description of the rocket servo system.

Fault ID		Fault description
Historical	F1	Increased clearance due to mechanical friction
	F2	Motor performance degradation, speed reduction
	F3	Increased current in channel A
	F4	Increased clearance and increased current in channel A
Unseen	F5*	Decreased output voltage of the filter

maintain relatively high CT as their subspace dimensions align with the number of training samples. The CT values of LPP-Bayesian and PCA-KNN are relatively small because they belong to linear model classification methods. According to Table 5, the CT of RD-KICA varies significantly for different faults. For example, diagnosing a sample of IDV(8) takes only 0.2145 ms, while IDV(4) requires 1.1625 ms. This discrepancy primarily arises from the diagnostic strategy described in Section 3.4. During the diagnosis process, if $C_{\text{new}} = 0$ or $C_{\text{new}} = 1$ in (33), the fault type can be determined directly; while if $C_{\text{new}} > 1$, BBCs needs to be activated for further diagnosis. The introduction of this extra step increases the overall CT. As evidenced by Table 5, the CT comparison reveals significant efficiency gains: the proposed RD-KICA achieves an average CT of 0.5398 ms, which represents approximately only one-third of the 1.4259 ms CT required by RD-KICA-without-FVS.

Combining the previous experimental results on diagnostics, we have the following findings: the proposed RD-KICA substantially improves the diagnostic accuracy of historical faults compared to the reconstruction-based RRFS, which indicates that increasing the magnitude as a fault attribute has a positive effect on the diagnostic accuracy. Furthermore, compared to the pattern recognition-based methods KSFDA, ENBC, and CNN-LSTM, the proposed RD-KICA possesses the ability to identify unseen faults. Additionally, combined with the data in Tables 4 and 5, the ablation experiments of RD-KICA and RD-KICA-without-FVS jointly verified that using FVS for undersampling can improve online computational efficiency without affecting diagnostic accuracy.

4.3 Online diagnosis on the rocket servo system

The above two experiments verify the proposed RD-KICA in terms of diagnostic accuracy and computing time. This subsection focuses on the online diagnosis procedure of the RD-KICA using a rocket servo system, including fault reconstruction, re-diagnosis using BBC and diagnosis decision-making. The experimental data came from a semi-physical simulation platform, as shown in Figure 6. This servo system includes dual-channel actuator data, such as displacement, velocity feedback, and nozzle swing angle, encompassing a total of 32 process variables. In this experiment, five fault modes were injected, as shown in Table 6. Notably, F4 is a composite of F1 and F3, so there is a significant overlap between them.

In this experiment, 500 fault samples were collected for each fault mode (F1–F4), with 300 samples from each mode utilized to extract fault subspaces, denoted as $\boldsymbol{\Xi}_1$, $\boldsymbol{\Xi}_2$, $\boldsymbol{\Xi}_3$ and $\boldsymbol{\Xi}_4$. The remaining 200 samples were used to evaluate the diagnostic performance. Additionally, 200 samples of F5* were collected to test the method's ability to identify unseen faults. The proposed RD-KICA method is used for online diagnosis, and the results of fault reconstruction are illustrated in Figure 7. The reconstruction statistics $I^2_{(4)}(\boldsymbol{k}^*_{\text{new}})$ of sample points 1–200 and 401–600 are less than the control limit I^2_{limit}, indicating that subspace $\boldsymbol{\Xi}_4$ mistakenly eliminated the fault effects of samples from F1 and F3. This result is because F4 (as a composite of F1 and F3) covers all fault information of them in $\boldsymbol{\Xi}_4$. In contrast, for sample points 601–800, only $\boldsymbol{\Xi}_4$ can effectively eliminate the fault effects, while $\boldsymbol{\Xi}_1$ and $\boldsymbol{\Xi}_3$ are failed. For the unseen fault sample points 801–1000, only one sample point has a $I^2_{(4)}(\boldsymbol{k}^*_{\text{new}})$ less than I^2_{limit}, while the others are above it, indicating that no subspace can completely cover the fault information of these samples. This reconstruction result is consistent with the fault injection situation in Table 6.

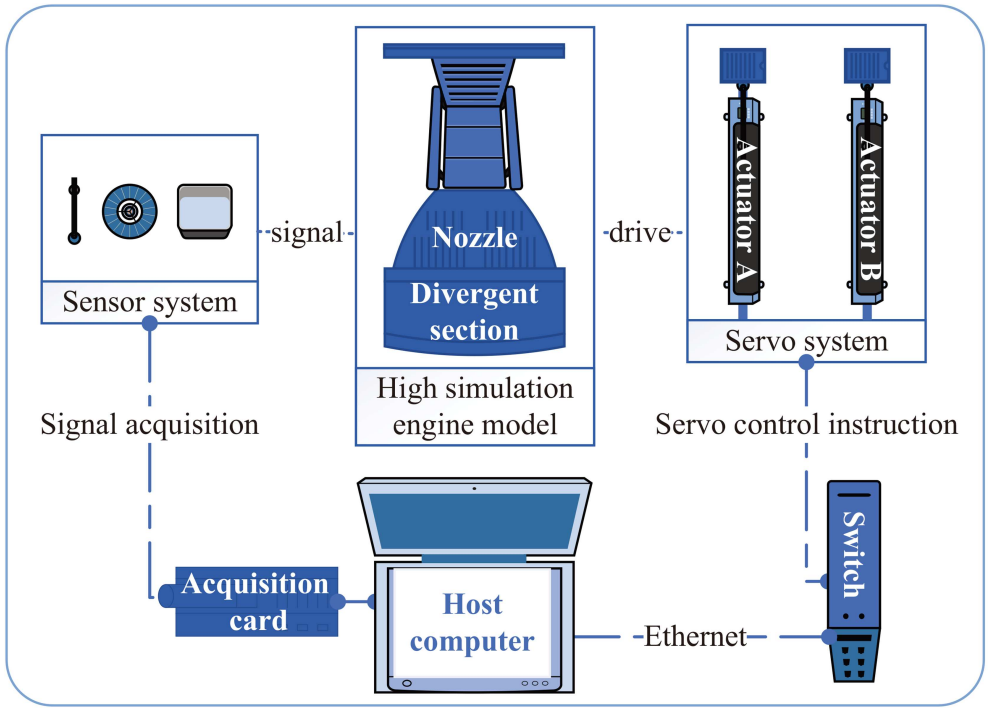

Figure 6 (Color online) Structure diagram of the rocket servo system.

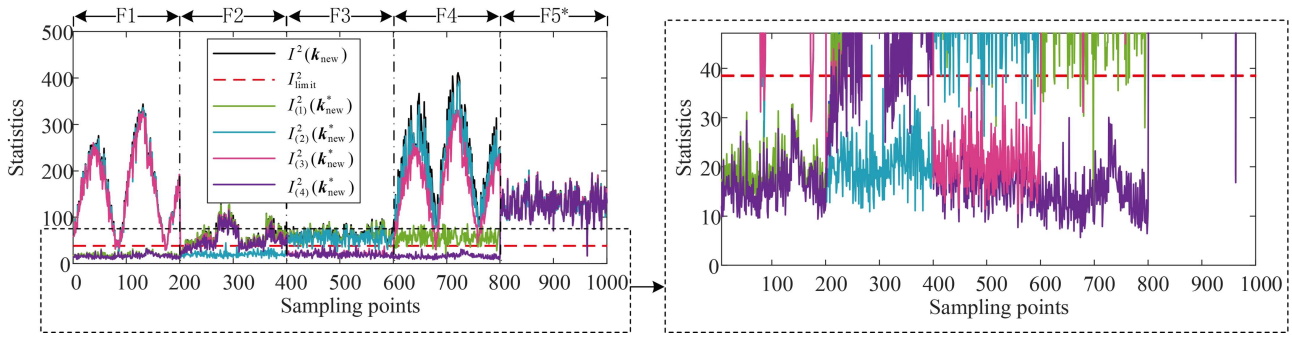

Figure 7 (Color online) Results of the fault reconstruction.

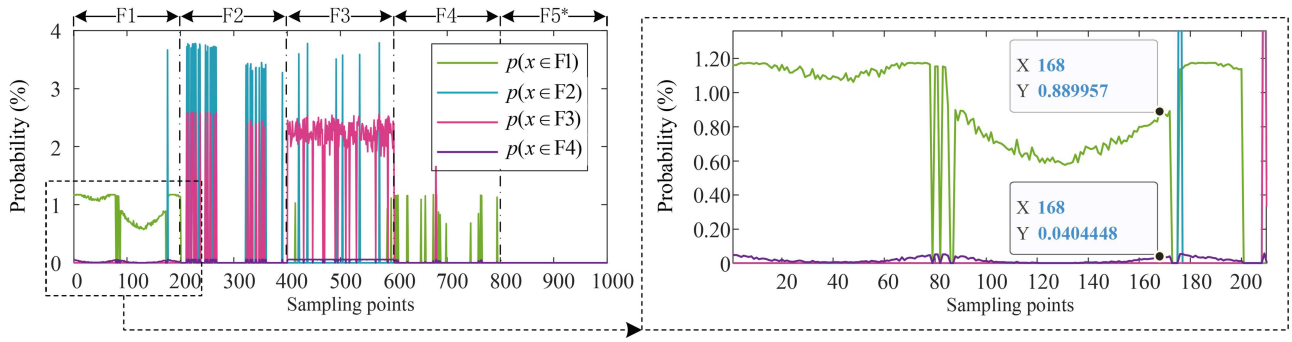

Figure 8 (Color online) Probabilities in re-diagnosis by BBC.

To address this incorrect elimination due to overlap, the proposed RD-KICA method incorporates a BBC for re-diagnosis. Figure 8 presents the re-diagnosis results, where sample points 1–200 are enlarged for clearer observation. The vertical axis of Figure 8 represents the probability of a sample belonging to a specific fault category. Next, we analyze this re-diagnosis using sample point 168 as an example. According to Figure 7, we can see that the reconstruction statistics $I^2_{(1)}(\boldsymbol{k}^*_{\text{new}})$ and $I^2_{(4)}(\boldsymbol{k}^*_{\text{new}})$ of this sample are less than I^2_{limit}. Therefore, we only calculate

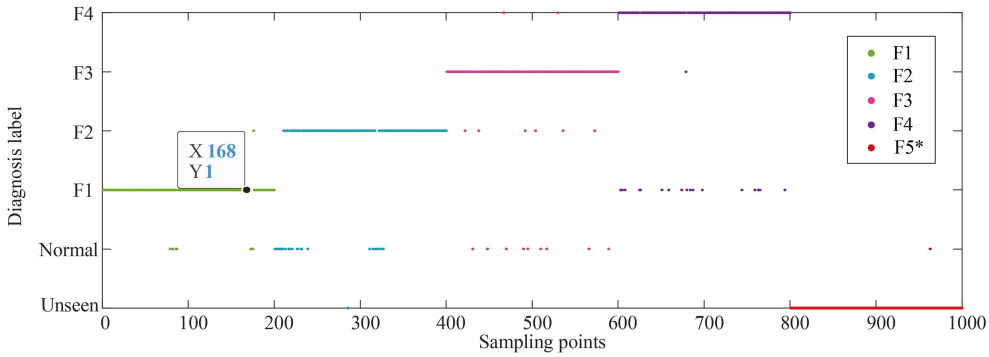

Figure 9 (Color online) Diagnostic results of the rocket servo system.

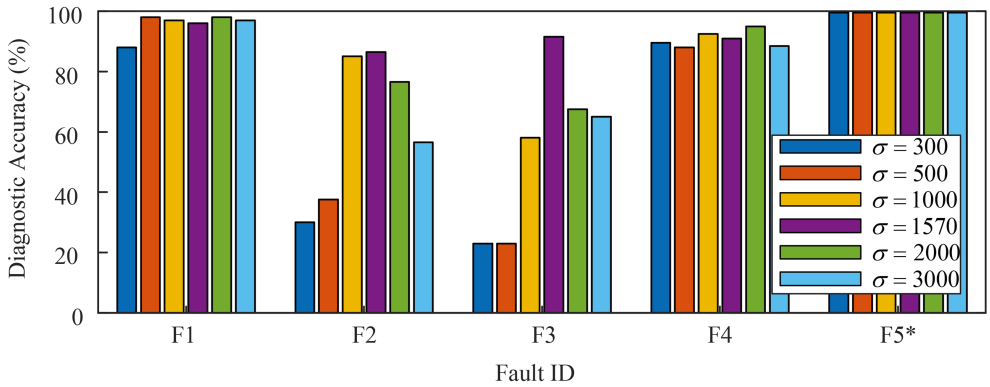

Figure 10 (Color online) Sensitivity of each fault to parameters σ in the rocket servo system.

the probabilities of this sample belonging to F1 and F4, which are 0.889957% and 0.0404448%, respectively. It means $p(\boldsymbol{x} \in \text{F1}) > p(\boldsymbol{x} \in \text{F4})$. Thus, we diagnose this sample as F1, as shown in Figure 9. It can be clearly observed from Figure 9 that most samples are accurately diagnosed, with only a few misdiagnoses. According to statistics, the diagnostic accuracy of the RD-KICA in this online diagnostic has reached 93.20%. Additionally, for sampling points of the unseen fault (i.e., sampling points 800–1000), except for one sample being misdiagnosed, the rest are correctly identified as unseen faults.

The key parameter of the proposed RD-KICA that influences model's performance is the kernel parameter σ. In this study, we employed a tabu search optimization for tuning [29, 30]. To investigate the effect of the kernel parameters on the model performance, we performed a sensitivity analysis by gradually varying the kernel parameters across a predefined range. For each parameter value, the fault diagnosis accuracy is calculated and summarized in Figure 10. Varying σ has a minimal effect on the diagnostic accuracy of F1, F4 and F5, while the diagnostic results of F2 and F3 show great fluctuation. Combined with the statistical analysis in Figure 7, it can be inferred that this fluctuation is mainly because the fault effect of F2 and F3 is not serious, and their statistics are only slightly higher than the control limit, resulting in some samples being misjudged as normal samples. The experimental results demonstrate that the kernel parameter σ significantly affects the diagnostic performance of RD-KICA. Notably, the kernel parameter optimized via tabu search achieves superior diagnostic accuracy.

5 Conclusion

In this paper, a reduced reconstruction modeling approach with dual attributes is proposed to tackle the high complexity and uncertainties during fault diagnosis. FVS undersampling is investigated to construct a less numerous but more informative training set. This approach is a novel attempt to combine the ideas of reconstruction and classification by taking subspaces and magnitudes together as dual attributes to build the fault library. Such that, the Bayesian classifier trained with magnitudes can continue to further diagnose under uncertainty when relying solely on fault subspaces. The main advantages of the RD-KICA include faster and more accurate diagnostic capabilities, as well as its ability to identify unseen faults. The proposed RD-KICA method is suitable for non-Gaussian nonlinear systems with multivariate variables and high system coupling, such as large complex equipment

systems, metal smelting, and intelligent manufacturing. It has unique advantages for systems with frequent unseen faults. The limitation of the proposed RD-KICA method lies in the requirement for structured fault sample data to establish the fault library. Issues such as multi-source heterogeneous data, missing values, or zero-shot learning scenarios still require further investigation and research.

This research has introduced a new research direction for fault attributes and re-diagnosis mechanisms. In this study, several Bayesian binary classifiers are used to identify the fault types; however, there are alternative methods that can be explored in future research, such as support vector machine (SVM), deep learning, and decision tree.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62273354, 62227814).

References

1 Sun C Y, Yang G H. A hierarchical quality-related fault diagnosis method for nonlinear industrial systems. IEEE Trans Ind Inf, 2023, 19: 8396–8405
2 Zhang N, Xu Y, Zhu Q X, et al. Farthest-nearest distance neighborhood and locality projections integrated with bootstrap for industrial process fault diagnosis. IEEE Trans Ind Inf, 2023, 19: 6284–6294
3 Li G, Joe Qin S, Zhou D. Output relevant fault reconstruction and fault subspace extraction in total projection to latent structures models. Ind Eng Chem Res, 2010, 49: 9175–9183
4 Zhao C, Sun Y. Subspace decomposition approach of fault deviations and its application to fault reconstruction. Control Eng Pract, 2013, 21: 1396–1409
5 Zhao C, Gao F. Fault subspace selection approach combined with analysis of relative changes for reconstruction modeling and multifault diagnosis. IEEE Trans Contr Syst Technol, 2016, 24: 928–939
6 Cai L, Tian X, Chen S. Monitoring nonlinear and non-Gaussian processes using Gaussian mixture model-based weighted kernel independent component analysis. IEEE Trans Neural Netw Learn Syst, 2017, 28: 122–135
7 Fang Z M, Zhang Y W, Deng R X, et al. Monitoring method of non-Gaussian process based on fractal analysis with kernel independent component regression. IEEE Trans Instrum Meas, 2023, 72: 3518809
8 Lee J, Qin S J, Lee I. Fault detection of non-linear processes using kernel independent component analysis. Can J Chem Eng, 2007, 85: 526–536
9 Liu Y, Wang F, Chang Y, et al. Performance-relevant kernel independent component analysis based operating performance assessment for nonlinear and non-Gaussian industrial processes. Chem Eng Sci, 2019, 209: 115167
10 Azim N M, Sarath R. Combined classification models for bearing fault diagnosis with improved ICA and MFCC feature set. Adv Eng Software, 2022, 173: 103249
11 Kong X Y, Yang Z Y, Luo J Y, et al. Extraction of reduced fault subspace based on KDICA and its application in fault diagnosis. IEEE Trans Instrum Meas, 2022, 71: 3505212
12 Jiang B, Zhu B. Dynamic Bhattacharyya bound-based approach for fault classification in industrial processes. IEEE Trans Ind Inf, 2022, 18: 397–404
13 Melani A H A, Michalski M A C, da Silva R F, et al. A framework to automate fault detection and diagnosis based on moving window principal component analysis and Bayesian network. Reliability Eng Syst Saf, 2021, 215: 107837
14 Yang G, Gu X. Fault diagnosis of complex chemical processes based on enhanced naive Bayesian method. IEEE Trans Instrum Meas, 2020, 69: 4649–4658
15 Tang Q, Li B Q, Chai Y, et al. Improved sparse representation based on local preserving projection for the fault diagnosis of multivariable system. Sci China Inf Sci, 2020, 64: 129204
16 Jiang B, Guo Z, Zhu Q, et al. Dynamic minimax probability machine-based approach for fault diagnosis using pairwise discriminate analysis. IEEE Trans Contr Syst Technol, 2019, 27: 806–813
17 Zhang N, Xu Y, Zhu Q X, et al. Improved locality preserving projections based on heat-kernel and cosine weights for fault classification in complex industrial processes. IEEE Trans Rel, 2023, 72: 204–213
18 Ge Z, Zhong S, Zhang Y. Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes. IEEE Trans Ind Inf, 2016, 12: 1403–1411
19 Han Y, Qi W, Ding N, et al. Short-time wavelet entropy integrating improved LSTM for fault diagnosis of modular multilevel converter. IEEE Trans Cybern, 2022, 52: 7504–7512
20 Huang D Q, Fu Y Z, Qin N, et al. Fault diagnosis of high-speed train bogie based on LSTM neural network. Sci China Inf Sci, 2021, 64: 119203
21 Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netws, 2000, 13: 411–430
22 Baudat G, Anouar F. Feature vector selection and projection using kernels. Neurocomputing, 2003, 55: 21–38
23 Vapnik V. Three remarks on the support vector method of function estimation. In: Advances in Kernel Methods. Cambridge: MIT Press, 1999. 25–41
24 Keller J Y, Darouach M, Krzakala G. Fault detection of multiple biases or process leaks in linear steady state systems. Comput Chem Eng, 1994, 18: 1001–1004
25 Liu J, Zio E. Feature vector regression with efficient hyperparameters tuning and geometric interpretation. Neurocomputing, 2016, 218: 411–422
26 Islam M M, Lee G, Hettiwatte S N. Application of Parzen Window estimation for incipient fault diagnosis in power transformers. High Voltage, 2018, 3: 303–309
27 Huang J, Yan X. Quality relevant and independent two block monitoring based on mutual information and KPCA. IEEE Trans Ind Electron, 2017, 64: 6518–6527
28 Hu C, Luo J, Kong X, et al. Orthogonal multi-block dynamic PLS for quality-related process monitoring. IEEE Trans Automat Sci Eng, 2024, 21: 3421–3434
29 Zuo X, Li B, Huang X, et al. Optimizing hospital emergency department layout via multiobjective tabu search. IEEE Trans Automat Sci Eng, 2019, 16: 1137–1147
30 Liu M, Kong X, Luo J, et al. Novel reduced kernel independent component analysis for process monitoring. Trans Inst Meas Control, 2024, 46: 1374–1387