Reconstruction-based stacked sparse auto-encoder for nonlinear industrial process fault diagnosis

Indexed by:
Web of Science Group

Qihang Weng[a], Shaojun Ren[a,*], Baoyu Zhu[a], Yinfeng Jin[b], Fengqi Si[a]

[a] School of Energy and Environment, Southeast University, China
[b] Nanjing Nari-Relays Electric Co., Ltd, China

Highlights

- A novel RBSSAE based fault diagnosis framework is developed for large-scale system.
- A generalized RB index calculation algorithm for deep neural networks is proposed.
- The search procedure of reconstruction directions is optimized using the SFFS method.
- The optimal reconstruction magnitudes are calculated by Stephenson iterative method.
- The proposed RBSSAE has outstanding diagnostic performance and computational efficiency.

Abstract

The reconstruction-based (RB) approach can effectively suppress the misdiagnosis problem due to the smearing effect in fault isolation. However, the current exploration of the RB approach for large-scale nonlinear systems is still limited. Therefore, this paper proposes a reliable and effective fault diagnosis method based on a reconstruction-based stacked sparse autoencoder (RBSSAE) for high-dimensional industrial systems. In RBSSAE, a reconstruction-based index achieved by the Steffensen iterative method is developed to check whether the given variable(s) are responsible for the faults efficiently. However, the number of possible faulty variable combinations grows exponentially with the system dimension or actual abnormal variables, causing an unbearable computational burden for variable combination optimization. Hence, the proposed RBSSAE utilizes a sequential floating forward selection approach to rapidly isolate the most decisive combination of fault variables, meeting a requirement of online fault diagnosis. Finally, the effectiveness of the RBSSAE is verified on a numerical example and a real industrial case. Comparisons with other state-of-the-art methods are also presented.

Keywords

process monitoring, fault diagnosis, stacked sparse auto-encoder, reconstruction-based method, sequential floating forward selection

1. Introduction

Early warning and fault isolation through fault detection and diagnosis (FDD) technology when faults occur is vital to strengthen modern industrial production safety and save production costs (1). The existing FDD methods can be mainly classified into mechanism-driven and data-driven methods. Unlike mechanism-driven methods, data-driven methods can obtain correlations between system parameters from historical process data without detailed system structural parameters and internal principles (2). With the rapid development of industrial production process automation and informatization technology, data-driven methods have received extensive attention and research (3). Multivariate statistical analysis (MSA) method, such as principal component analysis (PCA) (4,5) and partial least squares (PLS) (6,7), is one of the most typical data-driven methods that has been intensively studied and successfully applied in the field of FDD. Typically, the MSA methods utilize

(*) Corresponding author.
E-mail addresses: Q. Weng qhweng@seu.edu.cn, S. Ren (ORCID: 0000-0002-0013-4522) rsj@seu.edu.cn, B. Zhu baoyuzhu@seu.edu.cn, Y. Jin laurent769974864@163.com, F. Si fqsi@seu.edu.cn

statistical indexes, such as the squared prediction error (SPE) and Hoteling's T^2, to monitor the process operation state. Once the process is detected as faulty, the next step is to isolate the root sources for further guiding fault analysis. The contribution plot (CP) method is the most commonly used fault isolation method, which is simple to calculate but prone to the smearing effect, resulting in serious misdiagnosis, particularly for complex faults in large-scale systems (8).

The underlying reason for the smearing effect is that the abnormal variations inevitably affect the latent variable space rather than only existing in the residual space (9). As a result, the contribution of non-faulty variables is incorrectly exaggerated and exceeds the control threshold, leading to erroneous fault isolation results. Fortunately, the reconstruction-based (RB) approach is recognized as an effective method for suppressing the smearing effect in linear and nonlinear systems (10,11). This approach minimizes the reconstruction residuals by tuning the observed data actively with a given fault direction, avoiding the error transfer to latent variable space. When the minimized reconstruction residual is below the confidence limit, the given fault direction can be one correct path to restore the faulty data to normal condition (10). It is worth noting that the RB approach applied in PCA has guaranteed the accuracy of univariate fault isolation with large fault magnitudes by rigorous mathematical proofs (12,13). However, many fault directions exist to fulfill this requirement if the number of fault variables is not limited (14). To address this issue, the researchers usually add an artificial constraint that the variables involved in the fault direction are minimal to ensure the uniqueness of the isolation results (15).

Nevertheless, the RB approach has to evaluate all possible combinations of variables because the real fault direction is unknown, leading to a severe computational burden (16). For this problem, researchers have mainly focused on optimizing the variable combination search logic to improve computational efficiency (14,17–19). Owing to the high complexity of modern industry, the process variables are highly nonlinear and highly coupled with each other. It is well known that linear MSA methods cannot accurately represent highly nonlinear relationships in many modern industrial systems (20). Hence, many nonlinear MSA methods aided by kernel or neural network techniques are proposed to handle nonlinear data

(21,22). The neural network method, such as auto-encoder (AE) and auto-associative neural network (AANN), incorporates nonlinear activation layers to compress the observed data to the latent variable space and then reconstruct it to the original space (23). But inevitably, neural network techniques also suffer from the smearing effect (24).

To solve this problem, researchers have made many attempts to improve the performance of neural network methods. Pawlik et al. (25) combined the neural network method with a novel parameter named relative differences product of network statistics (rDPNS) for diagnosing machine misalignment and unbalance. Hallgrímsson et al. (15) optimize the AE using sparsity constraints to decrease the contribution of process variables unrelated to the faulty variables. The RB approach has also been applied to the neural network technique to address the smearing effect. Ren et al. (11,26) introduced the RB approach to AANN and input training neural network (ITNN), and the results showed that these improved methods could effectively inhibit the influence of the smearing effect. Tang et al. (16) extended the application of the RB approach to the variational AE model. These methods use the BAB method to optimize the search procedure for faulty variables. However, the BAB method is still inefficient when facing high-dimensional systems. To address this situation, Ren et al. further optimize the iterative calculation of reconstruction magnitude and reconstruction direction search (27). However, the network structure of the above RB approach-assisted neural network techniques is fixed five-layer, which are still lacking in feature extraction capability for complex variable correlation in the face of large-scale data in modern industry.

Deep learning has been successfully applied in process monitoring for large-scale industrial systems in recent years via its more complex network structure for extracting deep features (28). Li et al. (29) proposed an attention-based highway bidirectional long short-term memory (AHBi-LSTM) network for fault diagnosis based on the raw vibration signal. Karabacak et al. (30) applied the convolutional neural network (CNN) in thermal imaging from worm gearboxes; the results show that the method performed better than the vibra- and sound-based diagnosis methods. Although there has been much work applying deep learning to the FDD technology, the investigations on a generalized RB-assisted deep neural

network framework are still limited for multivariate fault isolation.

Therefore, this work aims to develop a reconstruction-based stacked sparse auto-encoder for effective and accurate fault diagnosis in high-dimensional nonlinear systems. The main contributions of this work can be summarized as follows: (1) the proposed RBSSAE extends the application of RB approach to deep neural networks, which can efficiently suppress the misdiagnosis problem caused by the smearing effect while extracting deeper features; (2) introducing the Sequential floating forward selection method to search for the most decisive combination of faulty variables efficiently; (3) using the Steffensen iterative method for approximating the nonlinear derivative equations to obtain the corresponding optimal fault magnitude, so as to obtain the RB index promptly and precisely with small computational effort. These contributions allow the RBSSAE to accurately isolate faulty variables within a short period after a fault occurs. In the end, a numerical example and a real industrial case are demonstrated to test and validate the effectiveness of the RBSSAE.

2. Material and methods

2.1. The basic stacked sparse auto-encoder, SSAE

2.1.1. Auto-encoder, AE

AE is an unsupervised learning neural network with a three-layer structure, which usually consists of an encoder and a decoder, as shown in Fig. 1. Given a data set $X = [x^{(1)}, x^{(2)}, \ldots, x^{(N)}] \in \mathbb{R}^{n \times N}$ containing N samples with n variables, where $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}]^T \in \mathbb{R}^n$, $(1 \leq i \leq N)$ denotes the ith sample. In AE, the encoder f^{enc} projects the input $x^{(i)}$ into the l-dimensional latent space and obtains the latent variables $h^{(i)} \in \mathbb{R}^{l \times 1}$, which is formulated as:

$$h^{(i)} = f^{enc}(x^{(i)}) = \sigma(W^{enc}x^{(i)} + b^{enc}), \quad (1)$$

where $W^{enc} \in \mathbb{R}^{l \times n}$ and $b^{enc} \in \mathbb{R}^{l \times 1}$ are the weight matrix and bias vector of the encoder, respectively. The activation function of the encoder is set as $\sigma(\tau) = \frac{1}{1+exp(-\tau)}$.

Through the decoder f^{dec}, $h^{(i)}$ is transformed back to the original space and reconstructs the model input $x^{(i)}$ as:

$$\hat{x}^{(i)} = f^{dec}(h^{(i)}) = \sigma(W^{dec}h^{(i)} + b^{dec}), \quad (2)$$

where $\hat{x}^{(i)}$ is the model output of the AE, $W^{dec} \in \mathbb{R}^{n \times l}$ and

$b^{dec} \in \mathbb{R}^{n \times 1}$ are the weight matrix and bias vector of the decoder, respectively. The activation function of the decoder is the same as that of the encoder.

The AE network is trained by updating the weight matrix and bias vector through error backpropagation to minimize the loss function, as shown in Eq. (3).

$$L_{AE} = \sum_{i=1}^{N} \|\hat{x}^{(i)} - x^{(i)}\|^2 + \lambda_s(\|W^{enc}\|^2 + \|W^{dec}\|^2), \quad (3)$$

where L_{AE} is the loss function of the AE network, $\|\cdot\|^2$ denotes L_2 norms. The left term is the reconstruction error, while the right term is the regularized network structure penalty to prevent network overfitting, whose weight is controlled by λ_s.
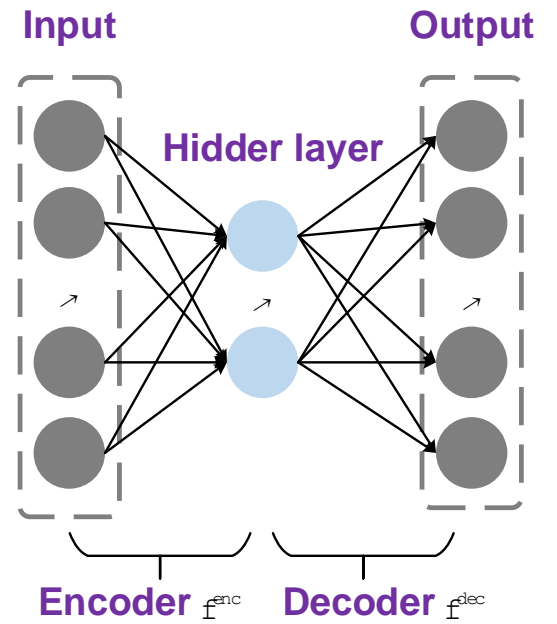

Fig. 1. Schematic diagram of the AE network.

2.1.2. Sparse auto-encoder, SAE

SAE improves the feature extraction efficiency by adding a sparsity constraint in the basic AE to control the average activation of the hidden layer neurons (31). The average activation measure of the hidden layer neuron j over the total sample set is given by:

$$\hat{\rho}_j = \frac{1}{N} \sum_{i=1}^{N} [\sigma_j(W^{dec}x^{(i)} + b^{enc})], \quad (4)$$

where $\sigma_j(\cdot)$ denotes the jth variable of the $\sigma(\cdot)$.

The relative entropy penalty term is added to the loss function to achieve the sparse constraint of the SAE, which is calculated as:

$$KL(\rho \parallel \hat{\rho}_j) = \log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}, \qquad (5)$$

where $KL(\cdot)$ denotes the Kullback-Leibler divergence. ρ is a predefined sparse target parameter.

The loss function of the SAE is given by:

$$L_{SAE} = L_{AE} + \lambda_{sparse}\sum_{j=1}^{l} KL(\rho \parallel \hat{\rho}_j), \qquad (6)$$

where λ_{sparse} controls the sparse penalty weight.

2.1.3. Stacked sparse auto-encoder, SSAE

SSAE is a deep neural network that arranges the base layers of multiple SAEs into a stacked structure through a fully connected form to obtain a more powerful nonlinear representation and more vital generalization ability. The SSAE stacked by m SAEs is shown in Fig. 2.
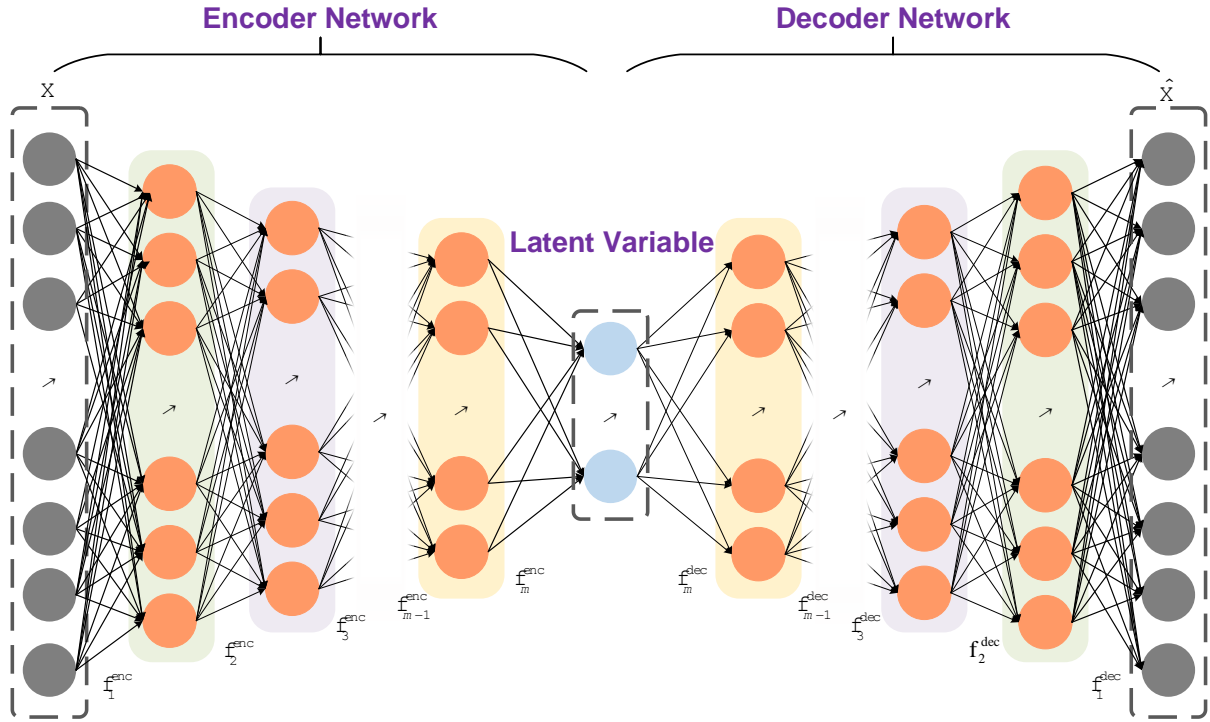

Fig. 2. Structure of the SSAE.

A layer-by-layer greedy unsupervised learning approach is used to optimize the inner parameter of SSAE. After training the outer SAE, their hidden layer is used as input to train the inner SAE, and finally, all the trained SAEs are cascaded to obtain the SSAE that has been parameter initialized. After performing pre-training, the parameters of each layer of the SSAE are locally optimal. The pre-trained SSAE is connected to the supervised classifier Softmax to fine-tune to obtain the optimal SSAE model. The Softmax cross-entropy loss function is:

$$L_{Softmax} = -\sum C_j \log\frac{e^{\hat{x}_j}}{\sum_k^K e^{\hat{x}_k}}, \qquad (7)$$

where C_j is the true classification label of the jth sample. K is the total number of categories. $\hat{x}_j$ is the jth input of the Softmax classifier.

Softmax uses the minimized cross-entropy loss to distinguish the correct or incorrect classification. Then, softmax adjusts the network parameters of the whole SSAE with the BP algorithm to make the parameter-adjusted classification labels of the integrated model as close to the true labels as possible, and finally achieves the purpose of optimizing the SSAE model.

2.2. Reconstruction-based sparse stacked autoencoder, RBSSAE

2.2.1. Reconstruction-based index

The Squared prediction error (SPE) statistic Q was used to monitor the residual space of the SSAE model, and the magnitude of its value reflects the degree of deviation between the observed data and model output,

$$Q(\boldsymbol{x}) = \|\tilde{\boldsymbol{x}}\|^2 = \|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2 = (\boldsymbol{x}-\hat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\hat{\boldsymbol{x}}), \qquad (8)$$

with a confidence limit:

$$\begin{cases} \delta_\alpha^2 = \theta_1 \left(1 - \dfrac{\theta_2 h_0 (1-h_0)}{\theta_1^2} + \dfrac{\zeta_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1}\right)^{\frac{1}{h_0}} \\ h_0 = 1 - \dfrac{2\theta_1 \theta_3}{3\theta_2^2} \\ \theta_i = \sum_{k=1}^{n} \lambda_k^i, i = 1,2,3 \end{cases}, \quad (9)$$

at $(1-\alpha) \times 100\%$ confidence level, where δ_α^2 is the threshold of $\mathcal{Q}$, and ζ_α is the standard score corresponding to $(1-\alpha)\%$. λ_k is the eigenvalues of the covariance matrix of $\widetilde{x}$. x is detected as an abnormal sample when $\mathcal{Q}(x) > \delta_\alpha^2$, otherwise, it will be classified as the normal sample.

Given an abnormal sample $x \in \mathbb{R}^{n \times 1}$, assume that the fault involved d variables $\Omega_i = \{x_{v_1}, x_{v_2}, \dots, x_{v_d}\}$, $d \leq n$. Then, the non-faulty sample z_{Ω_i} can be expressed as

$$z_{\Omega_i} = x - \Xi_{\Omega_i} f_{\Omega_i}, \quad (10)$$

where $\Xi_{\Omega_i} = [\xi_{v_1}, \xi_{v_2}, \dots, \xi_{v_d}] \in \mathbb{R}^{n \times d}$ denotes the hypothetical fault direction, where $\xi_{v_1} \in \mathbb{R}^{n \times 1}$ is the v_1-th standard orthonormal basis. For instance, $\xi_{v_1} = [0 \ \ 1 \ \ 0 \ \ 0 \ \ \dots \ \ 0]^{\mathrm{T}}$ denotes the hypothetical fault involved in the second variable x_2, $v_1 = 2$; $\xi_{v_2} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}^{\mathrm{T}}$ denotes the hypothetical fault involved in the second and first variables x_2 and x_1, $v_2 = [2 \ \ 1]$. $f_{\Omega_i} = [f_{v_1}, f_{v_2}, \dots, f_{v_d}]^{T} \in \mathbb{R}^{d \times 1}$ is the corresponding fault magnitude vector, f_{v_1} denotes the fault magnitude in x_{v_1}.

Denote the nonlinear mapping of the SSAE model as $\Phi(\cdot)$, SSAE model output of z_{Ω_i} is:

$$\hat{z}_{\Omega_i} = \Phi(z_{\Omega_i}) = f_1^{dec} \circ f_2^{dec} \circ \dots \circ f_m^{dec} \circ f_m^{enc} \circ f_{m-1}^{enc} \circ \dots \circ f_1^{enc}(z_{\Omega_i}), \quad (11)$$

where the symbol $\circ$ denotes the composite between the mappings. f_j^{enc} and f_j^{dec} denote the encoding and decoding process of the j-th SAE, respectively.

Then, the SPE index of z_{Ω_i} can be indicated as:

$$\mathcal{Q}(z_{\Omega_i}) = \|z_{\Omega_i} - \hat{z}_{\Omega_i}\|^2 = \|z_{\Omega_i} - \Phi(z_{\Omega_i})\|^2 = (z_{\Omega_i} - \hat{z}_{\Omega_i})^{\mathrm{T}}(z_{\Omega_i} - \hat{z}_{\Omega_i})$$
$$= (x - \Xi_{\Omega_i} f_{\Omega_i} - \hat{z}_{\Omega_i})^{\mathrm{T}}(x - \Xi_{\Omega_i} f_{\Omega_i} - \hat{z}_{\Omega_i})$$
$$= \|x - \hat{z}_{\Omega_i}\|^2 - 2(x - \hat{z}_{\Omega_i})^{\mathrm{T}}\Xi_{\Omega_i} f_{\Omega_i} + f_{\Omega_i}^{\mathrm{T}} f_{\Omega_i}. \quad (12)$$

Concerning $\mathcal{Q}(z_{\Omega_i})$, there is only one unknown parameter that is f_{Ω_i}. Hence, the next step is to find an optimized f_{Ω_i} that can obtain a minimum $\mathcal{Q}(z_{\Omega_i})$:

$$\mathcal{Q}_{\Omega_i}^{rb} = \min_{f_{\Omega_i}} \mathcal{Q}(z_{\Omega_i}) = \min_{f_{\Omega_i}} \|z_{\Omega_i} - \Phi(z_{\Omega_i})\|^2, \quad (13)$$

where $\mathcal{Q}_{\Omega_i}^{rb}$ is the reconstruction-based (RB) index along the fault direction Ξ_{Ω_i}. If the RB index is lower than the confidence limit $\mathcal{Q}_{\Omega_i}^{rb} < \delta_\alpha^2$, then the process can be brought back to normal along with the fault direction Ξ_{Ω_i}.

Researchers have given optimal existence proofs for RBPCA (32). However, this proof is not suitable for nonlinear systems such as SSAE. In order to obtain the corresponding reconstruction-based index, it is necessary to prove the existence of optimums applicable to nonlinear systems. For the convenience of the proof, the auxiliary function is first defined as:

$$\Psi: \mathbb{R}^n \mapsto \mathbb{R}, \Psi(f_{\Omega_i}) = \mathcal{Q}(z_{\Omega_i}) = \|x - \hat{z}_{\Omega_i}\|^2 - 2(x - \hat{z}_{\Omega_i})^{\mathrm{T}}\Xi_{\Omega_i} f_{\Omega_i} + f_{\Omega_i}^{\mathrm{T}} f_{\Omega_i}$$
$$= \|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|^2 - 2\left(x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\right)^{\mathrm{T}}\Xi_{\Omega_i} f_{\Omega_i} + \|f_{\Omega_i}\|^2. \quad (14)$$

As the sigmoid function is bounded, the following relation holds:

$$\|\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\| = \|\sigma(\tau_1^{dec})\| \leq \sqrt{n}, \quad (15)$$

where $\tau_1^{dec} = W_1^{dec}\left(f_2^{dec} \circ \dots \circ f_m^{dec} \circ f_m^{enc} \circ f_{m-1}^{enc} \circ \dots \circ f_1^{enc}(x - \Xi_{\Omega_i} f_{\Omega_i})\right) + b_1^{dec}$.

Consider the following limits:

$$\lim_{f_{\Omega_i} \to \infty} \frac{\Psi(f_{\Omega_i})}{\|f_{\Omega_i}\|^2} = \lim_{f_{\Omega_i} \to \infty} \frac{\|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|^2}{\|f_{\Omega_i}\|^2}$$
$$-2 \lim_{f_{\Omega_i} \to \infty} \frac{\left(x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\right)^{\mathrm{T}}\Xi_{\Omega_i} f_{\Omega_i}}{\|f_{\Omega_i}\|^2} + \lim_{f_{\Omega_i} \to \infty} \frac{\|f_{\Omega_i}\|^2}{\|f_{\Omega_i}\|^2}. \quad (16)$$

First, according to Eq. (15), it holds that $0 \leq \|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\| \leq \|x\| + \|\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\| \leq \|x\| + \sqrt{n}$, then $0 \leq \lim_{f_{\Omega_i} \to \infty} \frac{\|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|^2}{\|f_{\Omega_i}\|^2} \leq \lim_{f_{\Omega_i} \to \infty} \frac{(\|x\| + \sqrt{n})^2}{\|f_{\Omega_i}\|^2} = 0$, so $\lim_{f_{\Omega_i} \to \infty} \frac{\|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|^2}{\|f_{\Omega_i}\|^2} = 0$. Similarly, $0 = \lim_{f_{\Omega_i} \to \infty} \frac{-\|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|}{\|f_{\Omega_i}\|} = \lim_{f_{\Omega_i} \to \infty} \frac{-\|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|\|f_{\Omega_i}\|}{\|f_{\Omega_i}\|^2} \leq \lim_{f_{\Omega_i} \to \infty} \frac{\left(x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\right)^{\mathrm{T}}\Xi_{\Omega_i} f_{\Omega_i}}{\|f_{\Omega_i}\|^2} \leq \lim_{f_{\Omega_i} \to \infty} \frac{\|x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\|\|f_{\Omega_i}\|}{\|f_{\Omega_i}\|^2} \leq \lim_{f_{\Omega_i} \to \infty} \frac{\|x\| + \sqrt{n}}{\|f_{\Omega_i}\|} = 0$, so $\lim_{f_{\Omega_i} \to \infty} \frac{\left(x - \Phi(x - \Xi_{\Omega_i} f_{\Omega_i})\right)^{\mathrm{T}}\Xi_{\Omega_i} f_{\Omega_i}}{\|f_{\Omega_i}\|^2} = 0$.

Therefore, it is concluded that:

$$\lim_{f_{\Omega_i} \to \infty} \frac{\Psi(f_{\Omega_i})}{\|f_{\Omega_i}\|^2} = 0 - 0 + 1 = 1. \quad (17)$$

As a result, $\Psi(f_{\Omega_i}) \to +\infty$ when $f_{\Omega_i} \to \infty$. Then $\exists \delta > 0$, $\Psi(f_{\Omega_i}) > \Psi(\mathbf{0})$ when $\|f_{\Omega_i}\| > \delta$, where $\mathbf{0}$ denotes the zero vector. And since $\Psi(f_{\Omega_i})$ is continuous on $\overline{B_\delta(\mathbf{0})}$, where $\overline{B_\delta(\mathbf{0})}$ denotes the closed ball with center $\mathbf{0}$ and radius δ, there exists a minimum value of $\Psi(f_{\Omega_i})$ on $\overline{B_\delta(\mathbf{0})}$, written as ι. Clearly, $\iota < \Psi(f_{\Omega_i})$ when $f_{\Omega_i} \in \overline{B_\delta(\mathbf{0})}$, and $\iota < \Psi(\mathbf{0}) < \Psi(f_{\Omega_i})$ when $f_{\Omega_i} \in \mathbb{R}^d - \overline{B_\delta(\mathbf{0})}$. Therefore, ι is the minimum value of $\Psi(f_{\Omega_i})$ on $\mathbb{R}^d$, that is, there exists the minimum value of the nonlinear system $\mathcal{Q}(z_{\Omega_i})$.

Since the affine and activation layers of each SAE consist of smooth mappings, the nonlinear mapping of the SSAE model $\Phi(\cdot)$ is also smooth, so Eq. (13) can be solved by computing the zeros of the derivative equation of $\mathcal{Q}(z_{\Omega_i})$ concerning f_{Ω_i}, as follow:

$$\frac{d\mathcal{Q}(z_{\Omega_i})}{df_{\Omega_i}} = \frac{d\Psi(f_{\Omega_i})}{df_{\Omega_i}} =$$
$$2(\Phi(x - \Xi_{\Omega_i} f_{\Omega_i}) - x)^{\mathrm{T}} \frac{d\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})}{df_{\Omega_i}}$$
$$+ 2(\Xi_{\Omega_i} f_{\Omega_i})^{\mathrm{T}} \frac{d\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})}{df_{\Omega_i}} + 2f_{\Omega_i}, \quad (18)$$

with

$$\frac{d\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})}{df_{\Omega_i}} = (\prod_{k=1}^{m} Jf_k^{dec})(\prod_{k=1}^{m} Jf_{m-k+1}^{enc})(-\Xi_{\Omega_i}), \quad (19)$$

where Jf_k^{dec} denotes the Jacobian matrix of the mapping f_k^{dec}, Jf_k^{dec} and Jf_k^{enc} are given by:

$$Jf_k^{dec} = \mathrm{diag}\left[\sigma(\tau_k^{dec})\left(1 - \sigma(\tau_k^{dec})\right)\right] W_k^{dec}, \quad (20)$$

$$Jf_k^{enc} = \mathrm{diag}\left[\sigma(\tau_k^{enc})(1 - \sigma(\tau_k^{enc}))\right] W_k^{enc}. \quad (21)$$

Let $\frac{d\mathcal{Q}(z_{\Omega_i})}{df_{\Omega_i}} = 0$, the resulting $f_{\Omega_i}^*$ is the optimal fault magnitude corresponding to the hypothetical fault direction Ξ_{Ω_i}. At the end of the traversal of all hypothetical reconstruction directions, the final combination of the optimal reconstruction direction Ξ and magnitude f obtained that minimizes the final RB index of sample x:

$$\mathcal{Q}^{rb} = \min_{\Omega_i, \Omega_i \subset \Omega} \mathcal{Q}_{\Omega_i}^{rb}, \quad (22)$$

where Ω is the set containing all combinations of variables involved in the fault.

2.2.2. Steffensen method

In the traditional method of solving the RB index, there exists an analytical solution to the equation $\frac{d\mathcal{Q}(z_{\Omega_i})}{df_{\Omega_i}} = 0$. However, the model is strongly nonlinear for the SSAE-based residual reconstruction, and it is difficult to obtain the corresponding analytical solution. In this case, the equation $\frac{d\mathcal{Q}(z_{\Omega_i})}{df_{\Omega_i}} = 0$ is converted into a vector fixed point problem as shown in Eq. (23), then solved by the Steffensen method, which is faster than the Gradient descent (GD) and Momentum gradient descent (MGD) (27).

$$f_{\Omega_i} = \zeta(f_{\Omega_i}) = (\Phi(x - \Xi_{\Omega_i} f_{\Omega_i}) - x)^{\mathrm{T}} \frac{d\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})}{df_{\Omega_i}}$$
$$+ (\Phi(x - \Xi_{\Omega_i} f_{\Omega_i}) - x)^{\mathrm{T}} \Xi_{\Omega_i}$$
$$+ (\Xi_{\Omega_i} f_{\Omega_i})^{\mathrm{T}} \frac{d\Phi(x - \Xi_{\Omega_i} f_{\Omega_i})}{df_{\Omega_i}} + 2f_{\Omega_i}. \quad (23)$$

According to the Steffensen method, the iterative equation is obtained as follows:

$$f_{\Omega_i}^{(\tau+1)} = f_{\Omega_i}^{(\tau)} - [J^{(\tau)}]^{-1}[\zeta(f_{\Omega_i}^{(\tau)}) - f_{\Omega_i}^{(\tau)}], \quad (24)$$

where τ is the iteration counter, and J is the approximations of the Jacobian matrix, denoted as:

$$J = \begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1d} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{d1} & \omega_{d2} & \dots & \omega_{dd} \end{bmatrix}, \quad (25)$$

where ω_{jk} is the forward differential of the partial derivatives, calculated as:

$$\omega_{jk} = \frac{F_j[f_{v_1}^{(\tau)}, \dots, f_{v_k}^{(\tau)}, \zeta_{k+1}(f_{\Omega_i}^{(\tau)}), \dots, \zeta_d(f_{\Omega_i}^{(\tau)})] - F_j[f_{v_1}^{(\tau)}, \dots, f_{v_{k-1}}^{(\tau)}, F_k(f_{\Omega_i}^{(\tau)}), \dots, F_d(f_{\Omega_i}^{(\tau)})]}{F_j(f_{\Omega_i}^{(\tau)})},$$

$$(26)$$

with $F(f) = \frac{d\mathcal{Q}(x - \Xi_{\Omega_i} f)}{df}$.

The procedure for calculating the $\mathcal{Q}_{\Omega_i}^{rb}$ by Steffensen method is depicted in Algorithm 1.

Algorithm 1. $\mathcal{Q}_{\Omega_i}^{rb}$ calculation by the Steffensen method.

Step 1: Given an abnormal sample x, $\mathcal{Q}(x) > \delta_\alpha^2$. Initialize the iteration counter $\tau = 1$, the fault magnitude $f_{\Omega_i}^{(0)} = \mathbf{0}$, the convergence threshold $e = 0.0001$, the maximum number of iterations $\tau_{max} = 100$.

Step 2: Update $f_{\Omega_i}^{(\tau)}$ according to Eqs. (14)-(26) and then calculate $\mathcal{Q}_{\Omega_i}^{(\tau)}(z_{\Omega_i})$ by substituting $f_{\Omega_i}^{(\tau)}$ in Eq. (12).

Step 3: Update $\tau = \tau + 1$, and return to Step 2 until $\left| \mathcal{Q}_{\Omega_i}^{(\tau-1)}(z_{\Omega_i}) - \mathcal{Q}_{\Omega_i}^{(\tau)}(z_{\Omega_i}) \right| < e$ or $\tau \geq \tau_{max}$.

Step 4: $\boldsymbol{f}_{\Omega_i}^{(\tau)}$ is the optimized fault magnitude corresponding to $\boldsymbol{\Xi}_{\Omega_i}$, and $\mathcal{Q}_{\Omega_i}^{rb} = \mathcal{Q}_{\Omega_i}^{(\tau)}(z_{\Omega_i})$. End of the algorithm.

2.2.3. Variable selection via Sequential floating forward selection

The derivation of the RB index in the previous section is based on the case where the potential fault directions are known, except that the information about the actual fault directions is normally inadequate when a fault occurs. Therefore, locating the fault-related variables accurately and efficiently is another complex problem in the reconstruction method. To address this problem, the task of this section involves identifying d faulty variables from the n process variables that satisfy $Q^{rb} \leq \delta_\alpha^2$. It can be inferred from Proposition 1 that the more faulty variables are selected, the lower the RB index can be achieved.

Proposition 1. Consider a variable set $\mathcal{G}$ with d variables:
$$\forall v_k \in \mathcal{G}, Q_{\mathcal{G}-v_k}^{rb} \geq Q_{\mathcal{G}}^{rb}, \tag{27}$$
where $Q_{\mathcal{G}-v_k}^{rb}$ denotes the RB index along the reconstruction direction corresponding to $\mathcal{G}$ without v_k.

Proof: See Ref. (11).

It is necessary to restrict the number of faulty variables to avoid the fault directions tending to more faulty variables. The optimal fault directions should meet two conditions: (1) the RB index is smaller than the control threshold; (2) the RB index will be larger than the control threshold after removing any faulty variable from the fault directions, which can be represented as follow:
$$\begin{cases} Q_{\mathcal{G}}^{rb} < \delta_\alpha^2 \\ \forall v_k \in \mathcal{G}, Q_{\mathcal{G}-v_k}^{rb} > \delta_\alpha^2 \end{cases}. \tag{28}$$

Exhaustive search (ES) is a common and effective method for solving variable selection problems. However, the combinations of variables that need to be considered by ES are numerous under high dimensional systems with complex faults, which imposes a tremendous computational burden. On the other hand, the BAB method drastically reduces the amount of variable combinations to be retrieved by constructing a search tree compared to the ES method. However, there is still a severe computational boundary due to its search from high to low dimensional variable combinations, while the computation of

the RB index in the high dimensional reconstruction direction is costly and time-consuming. With these in mind, the SFFS, a multivariate suboptimal feature selection strategy based on a sequential greedy search algorithm, is applied to RBSSAE for variable selection. SFFS adds a deletion step on the basis of Sequential forward selection (SFS), which improves the disadvantages that SFS easily falls into local optimality and the redundancy of optimal feature sequences due to sequential greedy search. The SFFS framework for the RBSSAE is depicted in Algorithm 2.

Algorithm 2. Variable selection procedure via SFFS in RBSSAE.

Step 1: Given an abnormal sample $\boldsymbol{x}$, $\mathcal{Q}(\boldsymbol{x}) > \delta_\alpha^2$. Initialize the target set $\mathcal{G} = \emptyset$, the candidate set $\mathcal{F} = \{v_1, v_2, \cdots, v_n\}$ includes all variables.

Step 2: **Insertion**: seek the variable $v^* \in \mathcal{F}$ such that $v^* = \underset{v \in \mathcal{F}}{\arg\min} Q_{\mathcal{G} \cup v}^{rb}$, $Q_{\mathcal{G} \cup v}^{rb}$ is calculated according to Algorithm 1. Update $\mathcal{G} = \mathcal{G} \cup v^*$, $\mathcal{F} = \mathcal{F} - v^*$.

Step 3: Return to step 2 until $Q_{\mathcal{G}}^{rb} < \delta_\alpha^2$.

Step 4: **Deletion**: seek the variable $v^* \in \mathcal{G}$ such that $v^* = \underset{v \in \mathcal{G}}{\arg\min} Q_{\mathcal{G}-v}^{rb}$.

Step 5: Return to step 4, and update $\mathcal{G} = \mathcal{G} - v^*$ until $Q_{\mathcal{G}}^{rb} > \delta_\alpha^2$.

Step 6: The variables in $\mathcal{G}$ are isolated. End of the algorithm.

2.2.4. Fault diagnosis framework of RBSSAE

The fault diagnosis procedure of the RBSSAE method includes system modeling, fault detection, and fault isolation, as shown in Fig. 3. The detailed procedure for the RBSSAE is depicted in Algorithm 3.

Algorithm 3. Fault diagnosis procedure by the proposed RBSSAE.

Step 1: **System modeling**: develop a SSAE model by Section. 2 and obtain the confidence limit δ_α^2 by Eq. (9).
Step 2: Collect and initialize a new sample $\boldsymbol{x}$.

Step 3: **Fault detection**: calculate the $\mathcal{Q}(\boldsymbol{x})$ based on Eq. (8). Return to Step 2 until $\mathcal{Q}(\boldsymbol{x}) > \delta_\alpha^2$.

Step 4: **Fault isolation**: this process is detected as abnormal. Isolate the variables most responsible for the fault by Algorithm 2.
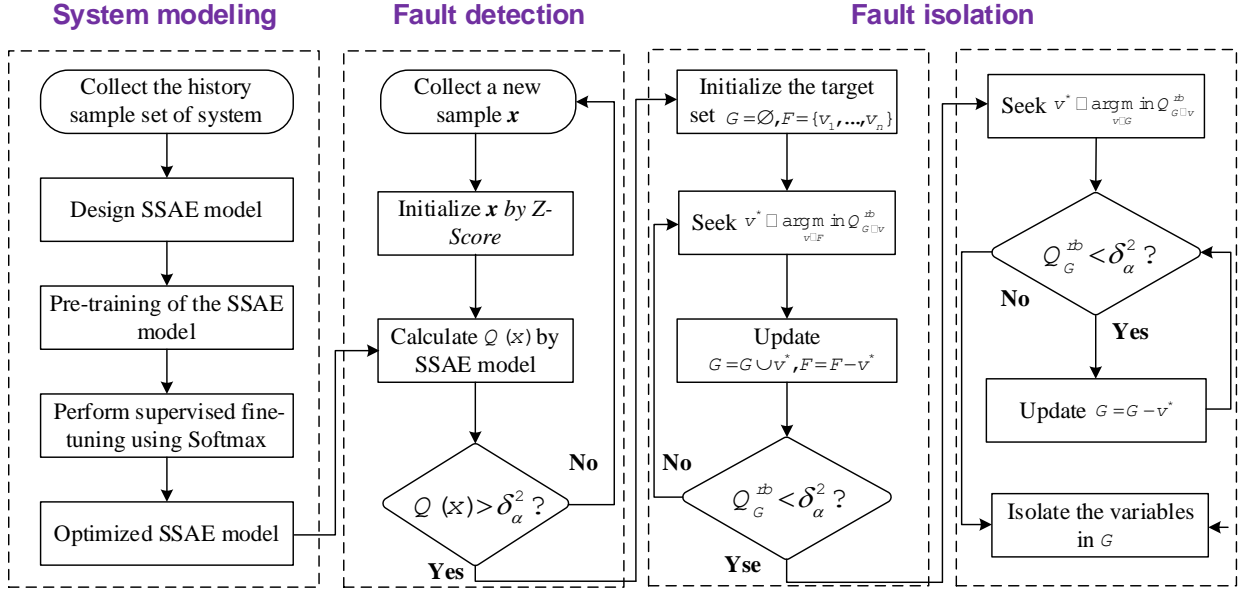
Step 5: End of the algorithm.

Fig. 3. The flowchart of the proposed RBSSAE for fault diagnosis.

3. Results and discussion

This section verifies the proposed RBSSAE method by a numerical example and a practical pulverizing system. All computation is carried out using MATLAB R2022b on a PC with Intel® Core™ i7–1260p CPU (2.10 GHz).

3.1. Numerical case

3.1.1. Data generation and fault simulation

The constructed 16-dimensional simulation process is generated by the following expressions:

$$
\begin{cases}
x_1 = v_1 + e_1 \\
x_2 = v_1 + 2 + e_2 \\
x_3 = 0.35(v_1^2 - 4v_1 - 1) + e_3 \\
x_4 = v_1^3 - 0.5 + e_4 \\
x_5 = 0.9x_3 + 0.15 + e_5 \\
x_6 = -x_4 + 0.25 + e_6 \\
x_7 = 1.2e^{-2v_1} + e_7 \\
x_8 = 0.18\cos(2\pi v_1) + v_1^2 + e_8 \\
x_9 = v_2 + e_9 \\
x_{10} = v_2 + 2 + e_{10} \\
x_{11} = 0.35(v_2^2 - 4v_1 - 1) + e_{11} \\
x_{12} = v_2^3 - 0.5 + e_{12} \\
x_{13} = 0.9x_{11} + 0.15 + e_{13} \\
x_{14} = -x_{12} + 0.25 + e_{14} \\
x_{15} = 1.2e^{-2v_2} + e_{15} \\
x_{16} = 0.18\cos(2\pi v_2) + v_2^2 + e_{16}
\end{cases}
\tag{29}
$$

where v_1 and v_2 are individually sampled from the uniform distribution $U[-1,1]$, respectively. $e_i, i = 1, \ldots, 16$ is a collection of independent noise variables following the normal distribution $N(0, 0.01)$.

First, 2000 samples are generated by Eq. (29), of which 1000 samples are used to train the model, and 1000 samples are used for the testing set. Then, five types of fault cases are simulated by the Monte Carlo method from the testing set, corresponding to faults involving one to five random variables. The generated fault magnitude follows a uniformly random distribution from -1 to 1.

To further explore the model's performance and fault diagnosis capabilities, four statistical indexes, including fault detection rate (FDR), false alarm rate (FAR), the absolute fraction of variance (R^2), and root mean square error (RMSE)), which are defined as follows:

$$
\text{FDR} = \frac{1}{n}\sum_{i=1}^{n}\frac{o_i}{r_i} \times 100\%,
\tag{30}
$$

$$
\text{FAR} = \frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{m - r_i} \times 100\%,
\tag{31}
$$

where n is the number of samples and m is the number of variables, o_i is the number of faulty variables isolated in the i th test sample, r_i is the total number of actual faulty variables in the i th test sample, p_i is the number of variables incorrectly isolated in the i th test sample.

$$
R^2 = 1 - \frac{\sum_{i=1}^{n}\left\|\hat{x}^{(i)} - x^{(i)}\right\|^2}{\sum_{i=1}^{n}\left\|x^{(i)} - \overline{x}\right\|^2},
\tag{32}
$$

where $\overline{x}$ is the mean value of all observed testing samples.

$$
\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{x}^{(i)} - x^{(i)}\right\|^2}.
\tag{33}
$$

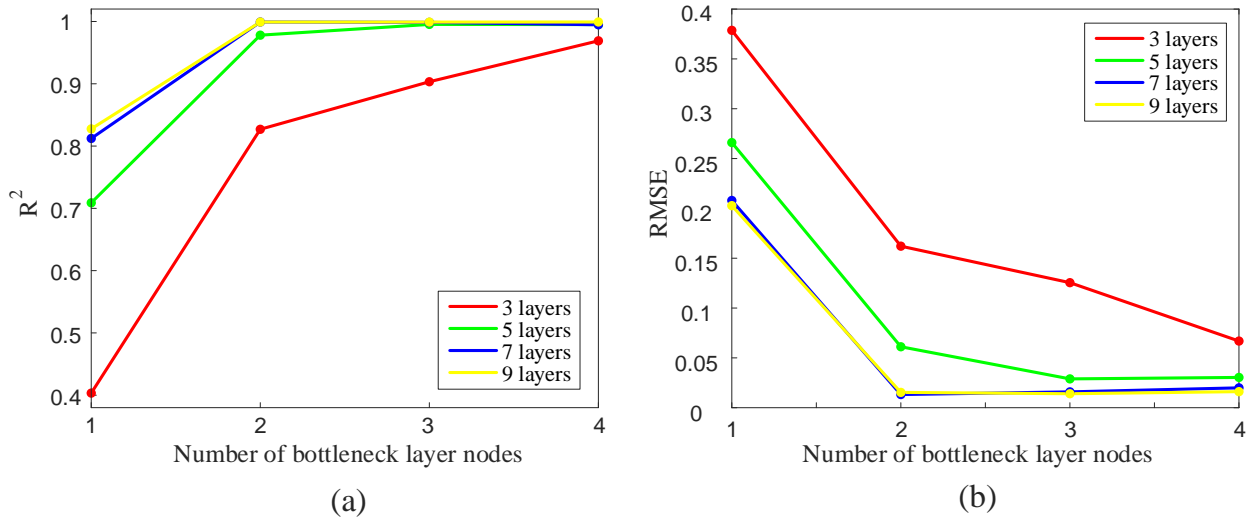### 3.1.2. Hyper-parametric sensitivity analysis

Fig. 4. Comparisons of (a) R^2 and (b) RMSE for different network structures.

The selection of hyperparameters significantly impacts the model training of auto-encoding networks. Fig. 4 illustrates the comparison of model performance under different network structures. It can be seen from the figure that with the increase in the number of layers of the network, the R^2 of the model gradually increases, and the RMSE gradually decreases. This indicates that the deeper auto-encoder network structure can improve the model's feature extraction ability to learn the correlation between the system parameters and achieve better model performance. Further, the auto-encoder model with a 7-layer network structure is significantly better than that of the 3-layer and 5-layer networks. However, it is not easy to obtain significant performance improvement by continuing to increase the network depth. Similarly, more bottleneck layer nodes can ease the difficulty of feature extraction during model learning, thus obtaining good model performance under the same number

of network layers. When the number of bottleneck layer nodes is 2, the model performance has reached the optimal value, and it is difficult to obtain significant model performance improvement when the number of bottleneck layer nodes continues to increase. Therefore, a 7-layer auto-encoder network and two bottleneck layer nodes were selected as the model structure.

3.1.3. Comparisons with other approaches

For comparison, five different methods are carried out in this section: (a) contribution plot-based PCA (CP-PCA), (b) reconstruction-based PCA (RBPCA), (c) contribution plot-based SSAE (CP-SSAE), (d) reconstruction-based SSAE aided by the ES search (RBSSAE-ES), and (d) the proposed reconstruction-based SSAE aided by the SFFS search (RBSSAE-SFFS).

Table 1. The fault diagnosis results of different methods.

Faulty variables	CP-PCA		RBPCA		CP-SSAE		RBSSAE-ES		RBSSAE -SFFS	
	FDR/%	FAR/%	FDR/%	FAR/%	FDR/%	FAR/%	FDR/%	FAR/%	FDR/%	FAR/%
1	36.893	1.066	33.500	0.167	91.000	14.133	92.500	0.033	**92.500**	**0.033**
2	34.500	2.300	35.750	0.179	89.750	24.786	93.250	0.036	**93.250**	**0.036**
3	31.636	3.145	34.333	0.462	89.667	31.038	93.000	0.077	**92.667**	**0.308**
4	28.966	3.580	40.375	0.417	88.875	39.833	93.000	0.042	**92.250**	**0.417**
5	24.280	4.569	38.700	0.727	91.100	44.318	92.500	0.136	**92.000**	**0.591**

Table 1 shows the fault diagnosis performance of different models in five types of fault cases, and the results of the proposed RBSSAE-SFFS are bolded. As can be seen from Table

1, the FDR of the two PCA-based methods is lower than 40% for each fault case since the PCA method cannot deal with nonlinear relationships. Moreover, the FAR of RBPCA is much

lower than CP-PCA. This is because the RB approach can effectively suppress the influence of the smearing effect compared to the CP method, thus reducing the generation of misdiagnosis situations. In contrast, several SSAE-based approaches can adapt well to the nonlinearity and accurately identify the faults with high FDRs around or above 90%. While observing the FAR results of CP-SSAE and RB-SSAE, the FAR of CP-SSAE increases dramatically with the increase of fault parameters, from 14.133% for single-parameter fault to 44.318% for five-parameter fault. The reason is that more faulty variables can significantly affect the faulty contribution of normal

variables, leading to faulty contribution of more normal variables being abnormally exaggerated and incorrectly exceeding thresholds. This results in worse misdiagnosis. The FAR of RBSSAE, on the other hand, always maintains a low value, which again leads to the conclusion that fault isolation by the RB approach can effectively suppress the influence of the smearing effect. Comparing the results of RBSSAE-ES and RBSSAE-SFFS, it can be observed that the FDR and FAR of RBSSAE-SFFS are slightly inferior to those of RBSSAE-ES. This is due to the influence of the SFFS search strategy, which will be discussed in more detail at the end of this section.
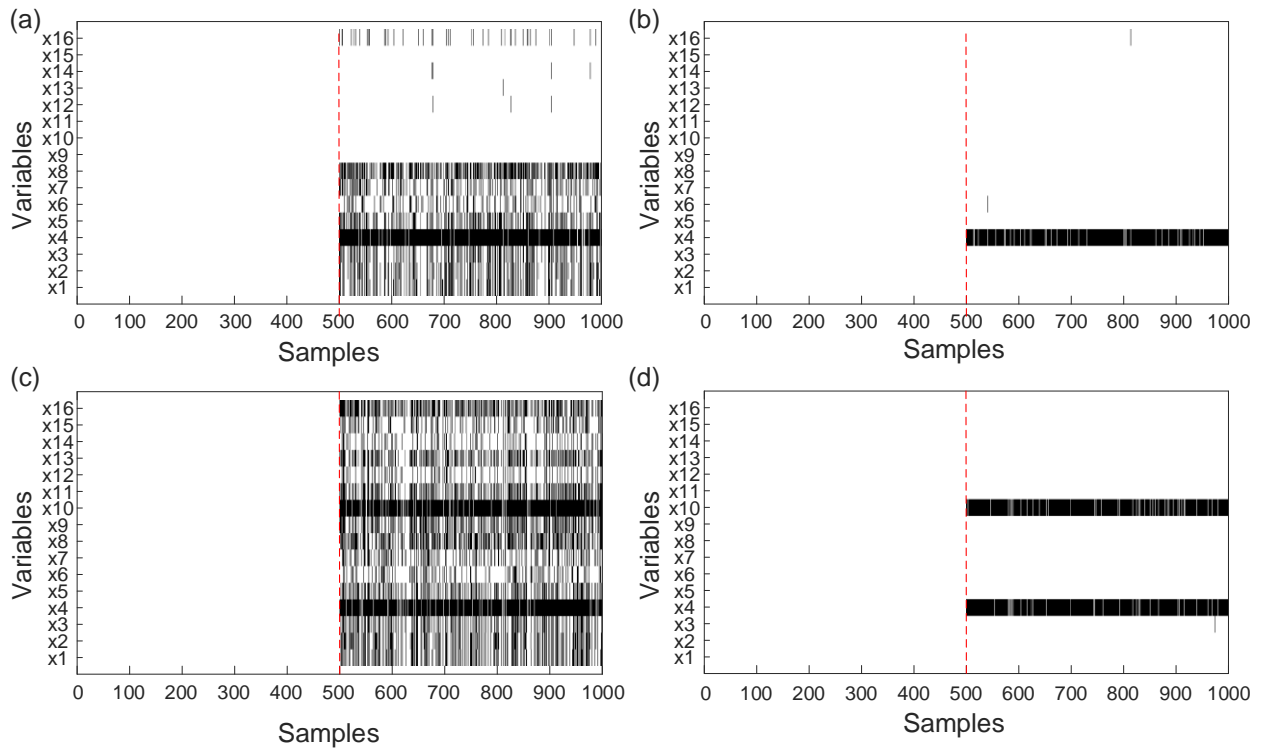

Fig. 5. The fault isolation results of one fault parameter by (a) CP-SSAE and (b) RBSSAE, two fault parameters by (c) CP-SSAE and (d) RBSSAE.

In order to compare the fault isolation effectiveness of the CP and RB approaches and the effect of smearing effect more visually, two testing sets of 1000 samples are generated: (1) the first 500 samples of two sets are normal data; (2) the second 500 samples simulate faults on the fourth variable and on both the fourth and tenth variables, respectively. The fault isolation results by CP-SSAE and RBSSAE are shown in Fig. 5, where the black cells represent the isolated variables, and the red line is the dividing line between normal and faulty samples. Fig. 5 demonstrates that the SSAE model can effectively discriminate between normal and abnormal samples. However, as can be seen from Fig. 5 (a) and (c), the CP method isolates many

normal variables in addition to the simulated faulty variables in the fault isolation stage, and as the fault parameters increase, more normal variables are incorrectly isolated. This is because the abnormal variable pulls up the faulty contribution of the remaining variables and even exceeds the control threshold, causing these normal variables to be incorrectly isolated by the CP method. In contrast, from Fig. 5 (b) and (d), the RB approach can accurately isolate the simulated faulty variables regardless of the single or multiple parameters fault. The reason is that the RB approach locates the faulty variables by the criterion of reconstructing the variables to pull the faulty samples back to the control thresholds without the influence on normal variables.
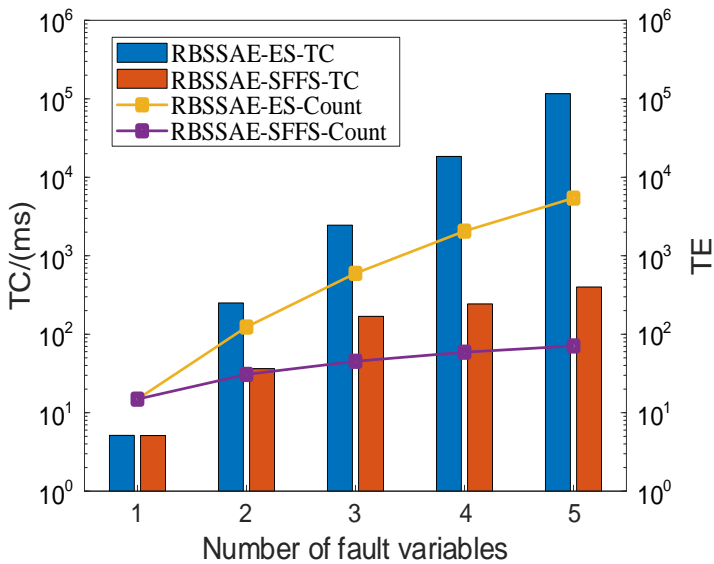
Fig. 6 Comparison of TC and TE for RBSSAE-ES and RBSSAE-SFFS.

The RB approach can obtain excellent fault isolation results, but it also brings a severe computational burden. In order to further demonstrate the superiority of the proposed method in terms of computational efficiency, a comparison is carried out between RBSSAE-ES and RBSSAE-SFFS. Fig. 6 illustrates the average time consumption (TC) and the average times of RB evaluations (TE) of one fault sample for the two methods. Fig. 6 shows that in the case of the faults involving a single variable, the TC and TE values by the ES and the SFFS are consistent since both methods need to traverse all variable combinations. As the number of faulty variables increases, the TC and TE for the ES search increase exponentially, whereas the SFFS search maintains a stable and controllable growth. Unlike the ES approach, the SFFS search is a greedy search method: when searching for the following variable, the variable combination of the previous search round is retained, equivalent to an univariate search. Instead, the ES search needs to re-traverse all possible variable combinations when performing variable combination element expansion. Such a search logic brings significant efficiency improvement to SSAE aided with the SFFS search, but it also carries the risk that the search falls into a local optimum. As can be seen from Table 1, RBSSAE-SFFS has different degrees of diagnostic degradation compared to RBSSAE-ES under 3, 4, and 5 parameters fault cases, which is caused by the local optimization problem of the SFFS search algorithm. Fortunately, the computational efficiency improvement is more objective than such a magnitude of

degradation. The proposed RBSSAE-SFFS has excellent fault diagnosis capability and is more competent to handle complex faults in large-scale systems.

3.2. The pulverizing system in coal-fired power plants

This section evaluates the proposed RBSSAE-SFFS method using three actual failure cases of the pulverizing system in a subcritical 600MW unit of a coal-fired power plant. The pulverizing system is equipped with a ZGM-113N medium-speed coal mill, the practical photos shown in Fig. 7.
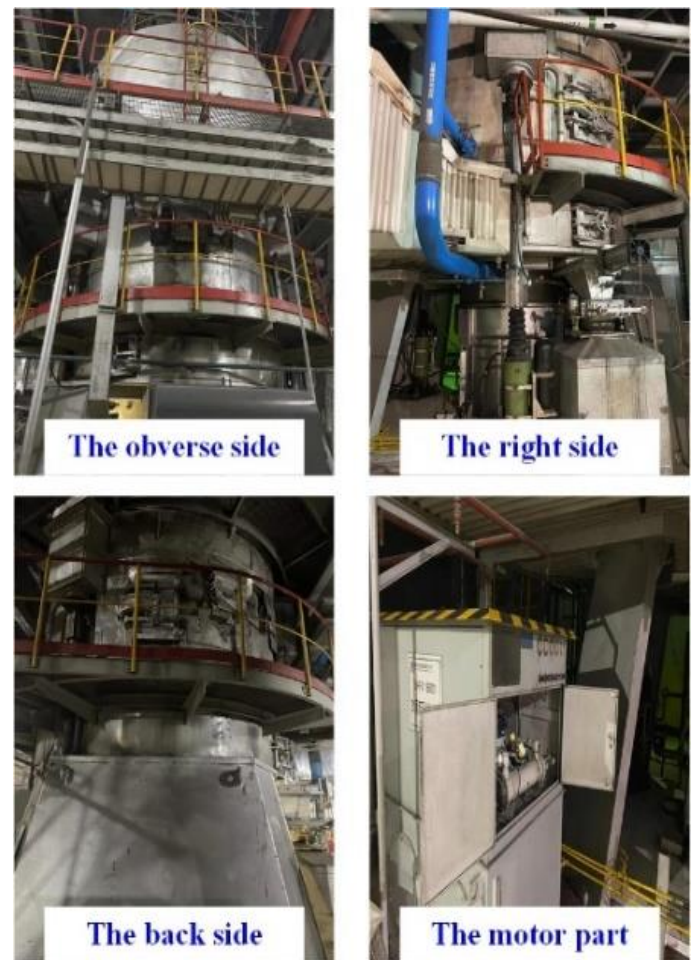

Fig. 7. The practical photos of the pulverizing system.

As shown in Fig. 8, the pulverizing system mainly consists of a raw coal bunker, coal feeder, coal mill, and seal fan. The raw coal is first stored in the raw coal bunker after entering the pulverizing system from the coal bin. Then, the raw coal is sent to the coal feeder, which feeds the raw coal evenly into the coal mill according to the required quantity. Next, the raw coal falls into the mill and is ground and squeezed into pulverized coal of the required fineness. The primary air dries the pulverized coal and delivers it to the furnace for combustion at the desired
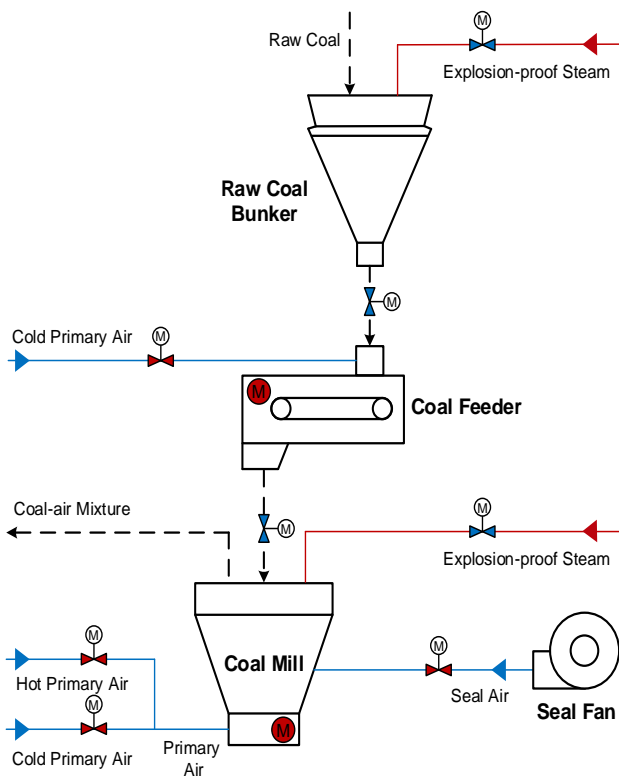
temperature.

Fig. 8. The Schematic diagram of the pulverizing system.

Table 2 lists 26 operating variables selected to build the pulverizing system fault diagnosis models. Ten days of operational data with 30-second intervals are collected from the plant-level Supervisory Information System (SIS) database. The coal type remained essentially constant during the sampling period. The samples are selected from different load segments by random non-repeated sampling to ensure the homogeneity of the training samples. Eventually, 8000 samples are selected to build a 26-20-14-6-14-20-26 SSAE model. Three real failure cases in system operation are collected to evaluate the validity of the proposed algorithm in the actual industrial process:

Case 1: A clogged sampling tube led to incorrect measurements of primary air pressure (M5).

Case 2: Abnormal measurements of the grinding roller bearing oil temperature (M14) caused by a damaged resistance temperature sensor.

Case 3: A coal blockage failure in the coal mill results in anomalous measurements of primary airflow rate (M2, M3), motor current of coal mill (M6), coal-air mixture outlet temperature (M8-M10), and differential pressure between inlet and outlet of coal mill (M24, M25).

Table 2. Operation intervals for selected variables.

No.	Variables	Description	Unit
M1	M_{coal}	Coal feeding capacity of coal feeder	t/h
M2-M3	W_{air}	Primary airflow rate	t/h
M4	T_{air}	Primary air temperature	°C
M5	P_{air}	Primary air pressure	kPa
M6	I_{mill}	The motor current of the coal mill	A
M7	I_{feed}	The motor current of the coal feeder	A
M8-M10	$T_{coal-air}$	Coal-air mixture outlet temperature (3)	°C
M11	$P_{coal-air}$	Coal-air mixture outlet pressure	kPa
M12	T_{tile}	Thrust tile temperature	°C
M13-M15	T_{oil}	The grinding roller bearing oil temperature (3)	°C
M16-M17	$T_{bearing}$	The motor bearing temperature (2)	°C
M18-M19	$T_{windingA}$	The motor electronic winding phase A temperature (3)	°C
M20-M21	$T_{windingB}$	The motor electronic winding phase B temperature (3)	°C
M22-M23	$T_{windingC}$	The motor electronic winding phase C temperature (3)	°C
M24-M25	ΔP_{in-out}	Differential pressure between inlet and outlet of coal mill (2)	kPa
M26	N_{unit}	Unit load	MW

Fig. 9 gives the fault isolation results by CP-SSAE and the proposed RBSSAE-SFFS in Case 1. Fig. 9 (a) shows the measured values of several sensors (M1, M5, M6, and M11), with the data scaled between 0 and 1 according to the maximum and minimum values for easy visualization. When the sampling tube clogs up, the primary air pressure (M5) measurement drops

rapidly and then keeps a lower value than the actual condition, as observed in Fig. 9 (a). Fig. 9 (b) demonstrates that the CP-SSAE exhibits good fault detection performance, can effectively identify the start and end of the fault, and can successfully isolate the faulty variable M5. However, a significant number of incorrect isolations occurred, especially

on M1, M6, and M11, due to the smearing effect. Since the faults did not involve these variables, as shown in Fig. 9 (a). Conversely, Fig. 9 (c) illustrates that RBSSAE-SFFS can accurately isolate the faulty variable M5 and suppress the misdiagnosis, reflecting better fault isolation performance than the CP-SSAE.
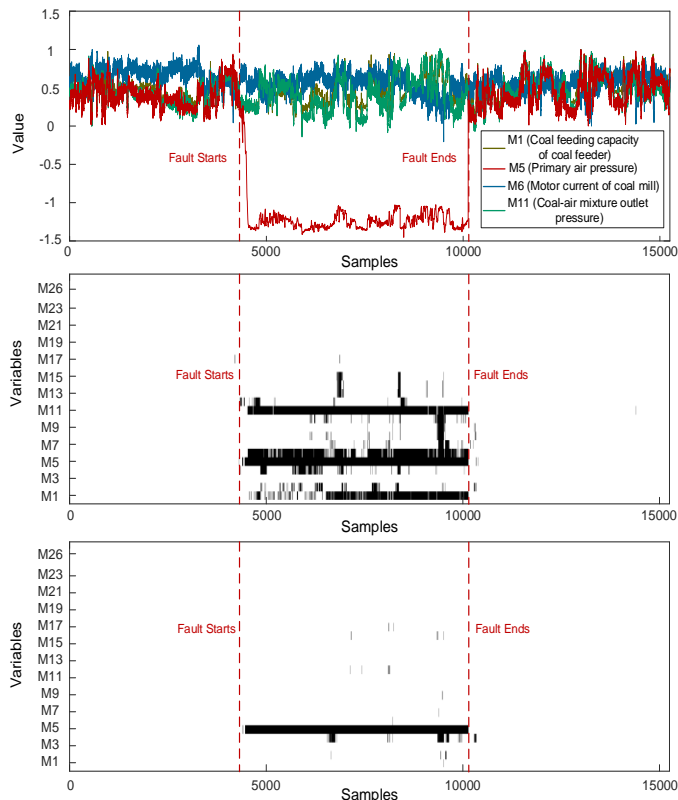

Fig. 9. The fault isolation results of Case 1: (a) the measured values of M1, M5, M6, and M11 with standardization, (b) fault isolation by the CP-SSAE, and (c) fault isolation by the RBSSAE-SFFS.

Fig. 10 displays the fault isolation results of Case 2 by two methods. As seen in Fig. 10 (a), the fault in Case 2 is not a step change fault as in Case 1. Instead, the magnitude of the fault in Case 2 becomes progressively more significant over time until the sensor is completely damaged. After replacing the sensor with a new one, the measured values returned to normal. Fig. 10 (b) and (c) indicate that while both methods can identify faults at an early failure stage, intermittent isolation results are observed due to the small magnitude. On the other hand, the alarms are promptly deactivated after replacement with new sensors, exhibiting excellent fault detection performance of the SSAE model. Additionally, unlike CP-SSAE, which generates many misdiagnosis cases, RBSSAE-SFFS can isolate the fault parameter efficiently and accurately.
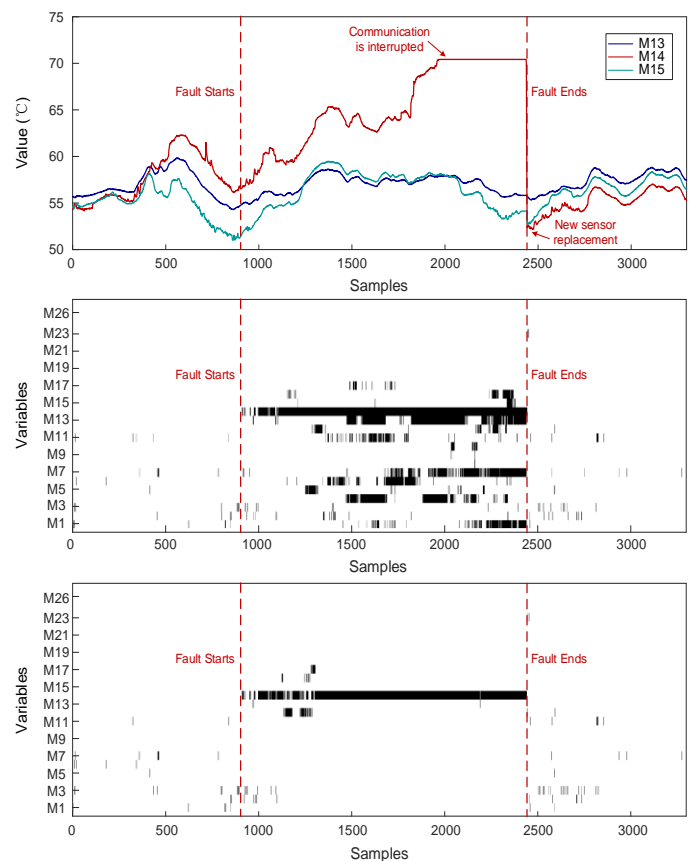

Fig. 10. The fault isolation results of Case 2: (a) the measured values of three grinding roller bearing oil temperature sensors M13, M14, and M15, (b) fault isolation by the CP-SSAE, (c) fault isolation by the RBSSAE-SFFS.

Fig. 11 illustrates the standardized measured values of the fault-involved variables and the fault isolation results of the two methods for Case 3. As shown in Fig. 11 (a), the coal mill current and the differential pressure between the inlet and outlet of the coal mill started to oscillate upward after the fault occurred. Besides, the primary airflow rate and the coal-air mixture outlet temperature also decreased continuously since the beginning of the fault. The coal blockage occurred at around the 190th sample. It was not detected and handled by the operator until the 350th sample, in other words, 80 minutes after the fault occurred, which severely disrupted the normal operations and production of the pulverizing system. Fig. 11 (b) and (c) demonstrate that the RBSSAE-SFFS can identify the fault at the early stage of the fault and accurately isolate the faulty variables to inform the following root cause troubleshooting. However, CP-SSAE results exhibit a more severe misdiagnosis caused by the smearing effect in complex fault cases compared to the single parameter faults in the previous two cases.
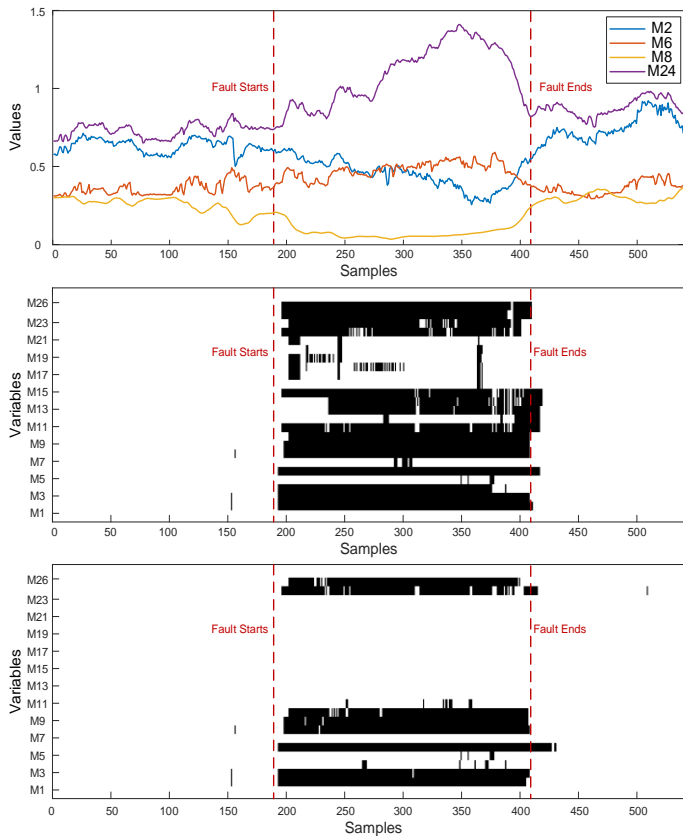
Fig. 11. The fault isolation results of Case 3: (a) the measured values of M2, M6, M8, M24, and M25 with standardization, (b) fault isolation by the CP-SSAE, and (c) fault isolation by the RBSSAE-SFFS.

Table 3. The fault diagnosis results of different methods.

Case	CP-PCA		RBPCA		CP-SSAE		RBSSAE-SFFS	
	FDR/%	FAR/%	FDR/%	FAR/%	FDR/%	FAR/%	FDR/%	FAR/%
1	99.982	5.995	99.789	0.316	99.982	7.355	99.982	0.248
2	64.128	5.655	76.432	2.248	95.117	10.141	94.892	0.346
3	18.838	2.104	13.283	1.970	89.040	33.235	91.893	1.133

Table 3 shows the fault diagnosis performance of different models in three real fault cases. As can be seen in Table 3, the results of Case 1 illustrate that the PCA-based method can still obtain good fault diagnosis results in nonlinear systems with simple faults, especially with large fault magnitudes. However, the results of Cases 2 and 3 show that it is difficult for PCA-based methods to effectively fault diagnose nonlinear systems under complex faults, especially with multiple fault variables. The SSAE-based method has a higher FDR in all three fault cases, which indicates that the SSAE-based method can effectively detect the occurrence of faults. CP-SSAE has a higher FAR than RBSSAE-SFFS, suggesting that CP-SSAE produces many misdiagnoses and proving the outstanding smearing effect contamination of the RB approach.

4. Conclusions

This paper proposes a new RBSSAE-based fault diagnosis method for large-scale nonlinear systems. This method employs the RB approach to isolate the faulty variables in the SSAE model. Meanwhile, it applies the SFFS method and the Steffensen iterative method to improve the computational efficiency of the RB approach. The numerical example demonstrates that RBSSAE can effectively detect and isolate high-dimensional nonlinear systems with high FDR and low FAR, regardless of simple or complex faults. Compared with RBPCA, CP-SSAE, and other RBSSAE, the proposed RBSSAE-SFFS shows excellent computational efficiency and fault diagnosis capability superiority. Moreover, a deeper network structure significantly improves the feature extraction capability of the model so that the deeper system features can be captured to achieve better model performance. Finally, the industrial case results show that RBSSAE-SFFS can effectively cope with actual industrial processes and overcome the smearing effects to ensure accurate fault isolation results.

Reference

1. Li Z, Bao S, Peng X, Luo L. Fault detection and diagnosis in multivariate systems using multiple correlation regression. Control Eng Pract [Internet]. 2021;116(January):104916. Available from: https://doi.org/10.1016/j.conengprac.2021.104916
2. Wen S, Zhang W, Sun Y, Li Z, Huang B, Bian S, et al. An enhanced principal component analysis method with Savitzky–Golay filter and clustering algorithm for sensor fault detection and diagnosis. Appl Energy [Internet]. 2023;337(July 2022):120862. Available from: https://doi.org/10.1016/j.apenergy.2023.120862
3. Kong X, Jiang X, Zhang B, Yuan J, Ge Z. Latent Variable Models in the Era of Industrial Big Data: Extension and Beyond. Annu Rev Control [Internet]. 2022;(June). Available from: https://doi.org/10.1016/j.arcontrol.2022.09.005
4. Zhang J, Zhou D, Chen M. Self-learning sparse PCA for multimode process monitoring. IEEE Trans Ind Informatics. 2022;XX(XX).

https://doi.org/10.1109/TII.2022.3178736

5. Wang Z, Zheng Y, Wong DSH. Trajectory-based operation monitoring of transition procedure in multimode process. J Process Control [Internet]. 2020;96:67–81. Available from: https://doi.org/10.1016/j.jprocont.2020.09.008

6. Kong XY, Cao ZH, Du BY, Luo JY. Quality-related multimodal fault detection technique based on partial least squares. Kongzhi yu Juece/Control Decis. 2019;34(12):2547–57.

7. Lin X, Sun R, Wang Y. Improved key performance indicator-partial least squares method for nonlinear process fault detection based on just-in-time learning. J Franklin Inst [Internet]. 2023;360(1):1–17. Available from: https://doi.org/10.1016/j.jfranklin.2022.11.029

8. Zheng Y, Zhou W, Yang W, Liu L, Liu Y, Zhang Y. Multivariate/minor fault diagnosis with severity level based on Bayesian decision theory and multidimensional RBC. J Process Control [Internet]. 2021;101:68–77. Available from: https://doi.org/10.1016/j.jprocont.2021.01.009

9. Van Den Kerkhof P, Vanlaer J, Gins G, Van Impe JFM. Analysis of smearing-out in contribution plot based fault isolation for Statistical Process Control. Chem Eng Sci. 2013;104:285–93. https://doi.org/10.1016/j.ces.2013.08.007

10. Yue HH, Qin SJ. Reconstruction-based fault identification using a combined index. Ind Eng Chem Res [Internet]. 2001;40(20):4403–14. Available from: https://doi.org/10.1021/ie000141+

11. Ren S, Si F, Zhou J, Qiao Z, Cheng Y. A new reconstruction-based auto-associative neural network for fault diagnosis in nonlinear systems. Chemom Intell Lab Syst. 2018;172(December):118–28. https://doi.org/10.1016/j.chemolab.2017.12.005

12. Mnassri B, El Adel EM, Ouladsine M. Reconstruction-based contribution approaches for improved fault diagnosis using principal component analysis. J Process Control [Internet]. 2015;33:60–76. Available from: http://dx.doi.org/10.1016/j.jprocont.2015.06.004

13. Alcala CF, Qin SJ. Reconstruction-based contribution for process monitoring. Automatica [Internet]. 2009;45(7):1593–600. Available from: http://dx.doi.org/10.1016/j.automatica.2009.02.027

14. Yan Z, Yao Y. Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO). Chemom Intell Lab Syst [Internet]. 2015;146:136–46. Available from: http://dx.doi.org/10.1016/j.chemolab.2015.05.019

15. Hallgrímsson ÁD, Niemann HH, Lind M. Unsupervised isolation of abnormal process variables using sparse autoencoders. J Process Control [Internet]. 2021;99:107–19. Available from: https://doi.org/10.1016/j.jprocont.2021.01.005

16. Tang P, Peng K, Jiao R. A process monitoring and fault isolation framework based on variational autoencoders and branch and bound method. J Franklin Inst [Internet]. 2022;359(2):1667–91. Available from: https://doi.org/10.1016/j.jfranklin.2021.11.016

17. Rauber TW, Boldt FA, Munaro CJ. Feature selection for multivariate contribution analysis in fault detection and isolation. J Franklin Inst. 2020;357(10):6294–320. https://doi.org/10.1016/j.jfranklin.2020.03.005

18. He B, Yang X, Chen T, Zhang J. Reconstruction-based multivariate contribution analysis for fault isolation: A branch and bound approach. J Process Control. 2012;22(7):1228–36. https://doi.org/10.1016/j.jprocont.2012.05.010

19. He B, Zhang J, Chen T, Yang X. Penalized reconstruction-based multivariate contribution analysis for fault isolation. Ind Eng Chem Res. 2013;52(23):7784–94. https://doi.org/10.1021/ie303225a

20. Zheng Y, Hu C, Wang X, Wu Z. Physics-informed recurrent neural network modeling for predictive control of nonlinear processes. J Process Control [Internet]. 2023;128:103005. Available from: https://doi.org/10.1016/j.jprocont.2023.103005

21. Alcala CF, Qin SJ. Reconstruction-based contribution for process monitoring with kernel principal component analysis. Proc 2010 Am Control Conf ACC 2010. 2010;7022–7. https://doi.org/10.1109/ACC.2010.5531315

22. Yu W, Zhao C. Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net. IEEE Trans Control Syst Technol. 2020;28(3):1083–91. https://doi.org/10.1109/TCST.2019.2897946

23. Li S, Luo J, Hu Y. Nonlinear process modeling via unidimensional convolutional neural networks with self-attention on global and local inter-variable structures and its application to process monitoring. ISA Trans [Internet]. 2022;121:105–18. Available from: https://doi.org/10.1016/j.isatra.2021.04.014

24. Hines JW, Wrest DJ, Uhrig RE. ERATION MONITO. JW Hines, DJ Wrest, RE Uhrig, Plant wide Sens calibration Monit Proc IEEE Internatinal Symp Intell Control IEEE, 1996, pp 378–383 . 1996;0–5.

25. Pawlik P, Kania K, Przysucha B. Eksploatacja i Niezawodnosc – Maintenance and Reliability neural network not requiring training data from a faulty machine. 2023;25(3):0–3. https://doi.org/10.17531/ein/168109

26. Ren S, Si F, Cao Y. Development of Input Training Neural Networks for Multiple Sensor Fault Isolation. IEEE Sens J. 2022;22(15):14997–

5009. https://doi.org/10.1109/JSEN.2022.3184078

27. Ren S, Jin Y, Zhao J, Cao Y, Si F. Nonlinear process monitoring based on generic reconstruction-based auto-associative neural network. J Franklin Inst [Internet]. 2023;360(7):5149–70. Available from: https://doi.org/10.1016/j.jfranklin.2023.03.041

28. Chen S, Yu J, Wang S. One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes. J Process Control [Internet]. 2020;87:54–67. Available from: https://doi.org/10.1016/j.jprocont.2020.01.004

29. Li X, Su K, He Q, Wang X, Xie Z. Research on Fault Diagnosis of Highway Bi-LSTM Based on Attention Mechanism Indexed by. Eksploat i Niezawodn. 2023;25(2):0–3. https://doi.org/10.17531/ein/162937

30. Özmen NG. Worm gear condition monitoring and fault detection from thermal images via deep learning method Monitorowanie stanu i wykrywanie błędów przekładni ślimakowej na podstawie termogramów z wykorzystaniem metody głębokiego uczenia. 2020;22(3):544–56. https://doi.org/10.17531/ein.2020.3.18

31. Liu J, Wu J, Xie Y, Jie W, Xu P, Tang Z, et al. Toward robust process monitoring of complex process industries based on denoising sparse auto-encoder. J Ind Inf Integr [Internet]. 2022;30(November):100410. Available from: https://doi.org/10.1016/j.jii.2022.100410

32. Dunia R, Joe Qin S. A unified geometric approach to process and sensor fault identification and reconstruction: The unidimensional fault case. Comput Chem Eng. 1998;22(7–8):927–43. https://doi.org/10.1016/S0098-1354(97)00277-9