Chinese Journal of Chemical Engineering 84 (2025) 146–157

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

SmdaNet: A hierarchical hard sample mining and domain adaptation neural network for fault diagnosis in industrial process

Zhenhua Yu, Zongyu Yao, Weijun Wang, Qingchao Jiang*, Zhixing Cao*

Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Article history:
Received 30 December 2024
Received in revised form
8 May 2025
Accepted 8 May 2025
Available online 22 May 2025

Keywords:
Industrial process
Bioprocess
Fault diagnosis
Neural networks
Fermentation

ABSTRACT

Fault diagnosis in industrial process is essential for ensuring production safety and efficiency. However, existing methods exhibit limited capability in recognizing hard samples and struggle to maintain consistency in feature distributions across domains, resulting in suboptimal performance and robustness. Therefore, this paper proposes a fault diagnosis neural network for hard sample mining and domain adaptive (SmdaNet). First, the method uses deep belief networks (DBN) to build a diagnostic model. Hard samples are mined based on the loss values, dividing the data set into hard and easy samples. Second, elastic weight consolidation (EWC) is used to train the model on hard samples, effectively preventing information forgetting. Finally, the feature space domain adaptation is introduced to optimize the feature space by minimizing the Kullback–Leibler divergence of the feature distributions. Experimental results show that the proposed SmdaNet method outperforms existing approaches in terms of classification accuracy, robustness and interpretability on the penicillin simulation and Tennessee Eastman process datasets.

1. Introduction

With the increasing complexity of industrial processes and advances in equipment intelligence, fault diagnosis has become critical to modern industrial production. Faults can significantly impact production efficiency and product quality while incurring substantial economic losses. Therefore, accurate and robust fault identification is essential. Fault diagnosis methods can be broadly classified into three types: physical model-based, domain knowledge-based and data-driven approaches [1–3].

Physical model-based methods represent processes through mathematical models that describe input-output relationships based on physical principles and laws. Common techniques include bond graphs [4], parameter estimation [5], Kalman filtering [6], state observers [7]. While accurate physical models can achieve high fault identification accuracy under ideal conditions, the increasing process complexity in industrial scenarios makes the construction of such models challenging and reduces their effectiveness in fault diagnosis. As a result, the use of physical model-based methods in industrial process fault diagnosis is declining.

Domain knowledge-based methods rely heavily on the expertise of engineers and are well suited for systems with sophisticated modelling mechanisms or insufficient sensor data. For example, fault tree analysis employs a tree diagram to systematically refine and prioritize fault causes [8]. Additionally, directed graphs and case-based reasoning are also widely used methods [9,10]. However, these methods are often complex and time-consuming to implement, resulting in high development costs and limited applicability.

Statistical analysis-based methods use statistical principles to analyze and model system operation data for fault diagnosis. Common methods include principal component analysis (PCA) [11], partial least squares (PLS) [12], and independent component analysis (ICA) [13]. Choi *et al.* [14] enhanced PCA by incorporating a kernel function, enabling efficient handling of nonlinear features and introducing a unified fault detection metric to improve sensitivity and accuracy. Lee *et al.* [15] proposed kernel independent component analysis (KICA), which improves the extraction of underlying data structures through high-dimensional mapping and independent component analysis. Yin *et al.* [16] unveiled the intrinsic relationship between process variables. Zhong *et al.* [17] proposed sparse local fisher discriminant analysis (SLFDA), which

This article is part of a special issue entitled: AI for Chemical Engineering published in Chinese Journal of Chemical Engineering.

* Corresponding authors.
E-mail addresses: qchjiang@ecust.edu.cn (Q. Jiang), zcao@ecust.edu.cn (Z. Cao).

https://doi.org/10.1016/j.cjche.2025.05.003

improves fault mode discrimination by incorporating sparsity constraints and local structural information.

Machine learning is a key approach in data-driven fault diagnosis and has been extensively applied. Common methods include artificial neural networks (ANNs) [18], K-nearest neighbors (KNN) [19], and support vector machines (SVM) [20]. ANNs identify fault types by adjusting node relationships, making them suitable for multi-class faults, but their parameter count increases significantly with larger datasets [18]. KNN classifies faults based on neighborhood samples. While simple to implement, it incurs high computational costs for global searches [19]. SVM performs fault diagnosis by constructing hyperplanes with strong generalization capabilities, making it especially effective on small datasets [20]. Although machine learning algorithms are effective for fault diagnosis, they are shallow models that cannot fully capture deep features.

Deep learning-based methods offer efficient and accurate solutions for fault diagnosis. Digital twin-assisted models incorporating multiscale residual and attention mechanisms have demonstrated strong capabilities in accurately identifying faults in complex systems such as hypersonic flight vehicles [21]. In parallel, task-oriented meta-learning techniques with gradient calibration have been introduced to enhance few-shot diagnostic generalization in fine-grained fault scenarios [22]. Additionally, recent studies have proposed dynamic domain adaptation frameworks to enable structural flexibility and multi-source knowledge alignment for fault diagnosis without requiring labeled target data [23]. Zhong *et al.* [24] proposed a hybrid fault diagnosis method that integrates convolutional neural network (CNN) and SVM. The CNN extracts features from raw data and transforms them into high-level representations, and SVM classify faults for effective state identification. To address the unstable feature distribution, Zhao *et al.* [25] used batch normalization to stabilize the data distribution and mitigate gradient vanishing and explosion. In addition, exponential moving averages were used to smooth parameter updates, improving model stability and generalization. Zhang *et al.* [26] proposed the wide kernel deep convolutional neural network (WDCNN), which extends the receptive field by widening the convolutional kernel, enabling efficient data structure and pattern capture while enhancing classification performance. Zhao *et al.* [27] introduced an LSTM-based recurrent neural network (RNN) for fault diagnosis, leveraging its capability to model long and short term dependencies. Zhang *et al.* [28] developed a fault diagnosis model using deep graph convolutional network (DGCN), which converts vibration signals into graph-structured data via wavelet transform and Pearson correlation. Zhou *et al.* [29] used deep convolutional generative adversarial networks (DCGAN) to achieve high precision fault diagnosis, effectively addressing the challenge of scarce labeled data. San *et al.* [30] proposed a time-aware variational autoencoder (tVAE) to significantly reduce the amount of labeled data required for fault diagnosis.

In recent years, deep belief networks (DBNs) have been extensively used to fault diagnosis, with research primarily focusing on parameter optimization, signal processing, and structural improvements. Wang *et al.* [31] used particle swarm optimization (PSO) to optimize the structure of DBNs. Deng *et al.* [32] applied an improved quantum-inspired differential evolutionary algorithm, enhancing the global search capability for parameter optimization. For signal processing, Chen *et al.* [33] integrated multi-sensor features using a sparse autoencoder; Jung *et al.* [34] employed multiscale vibration signal features as inputs; and Yan *et al.* [35] fused heterogeneous information to enhance rotor unbalance fault identification. In structural improvements, Niu *et al.* [36] introduced a parametric ReLU (PReLU) activation layer into DBNs (PReLU-DBN) to enhance convergence speed and diagnostic accuracy. However, these methods fail to address the impact of hard-to-differentiate samples on fault diagnosis models, leading to reduced classification accuracy, unstable decision boundaries, overfitting, and compromised robustness and practical effectiveness.

This paper proposes an improved DBN with hard sample mining and domain-adaptive for fault diagnosis (SmdaNet). It first constructs a fault diagnosis model based on DBN, dividing training data into hard and easy samples based on loss values. The DBN is then fine-tuned on the hard samples using an elastic weight consolidation (EWC), which updates only low-importance parameters to preserve knowledge from easy samples. Then, it optimizes feature representation by minimizing the Kullback–Leibler (KL) divergence between feature distributions of hard and easy samples. For model interpretability, t-SNE is used to visualize the features in the penultimate layer of the network. The contributions of this paper are as follows.

(1) An improved DBN integrating hard sample mining and EWC is proposed for industrial process fault diagnosis. It can preserve knowledge from easy samples while enhancing the classification accuracy of hard samples.
(2) A domain adaptation (DA) mechanism is introduced to improve cross-domain feature consistency, enhancing the model's discriminative power and interpretability.
(3) Experimental results on the penicillin simulation dataset and the Tennessee-Eastman (TE) Process dataset demonstrate that the proposed method outperforms existing fault diagnosis models.

The remainder of this article is structured as follows: Section 2 discusses the preliminaries. Section 3 provides a detailed explanation of the proposed SmdaNet algorithm. Section 4 presents the experimental results and analysis. Section 5 concludes the paper with a summary.

2. Preliminaries

2.1. DBN

A DBN consists of a stack of multilayer restricted Boltzmann machine (RBM), as shown in Fig. 1.

A RBM is composed of m observable variable and n hidden variable, and contains observable layer neurons $\boldsymbol{v} = \{v_1, v_2, \cdots, v_m\} \in [0, 1]$ and hidden layer neurons $\boldsymbol{h} = \{h_1, h_2, \cdots, h_n\} \in [0, 1]$. Given the weights w, the bias of the observable layer neurons a and the bias of the hidden layer neurons b, the energy function of a RBM is defined as:

$$E(v, h) = -\sum_{i}^{m} a_i v_i - \sum_{j}^{n} b_j h_j - \sum_{i}^{m}\sum_{j}^{n} v_i w_{ij} h_j \tag{1}$$

According to the energy function, the joint probability distribution $p(v, h|\theta)$ can be defined as

$$p(v, h|\theta) = \frac{1}{Z(\theta)} e^{-E(v,h)} \tag{2}$$

where $Z(\theta) = \sum_{v,h} e^{-E(v,h)}$ is the normalization factor. The edge probability distributions of the visible and hidden layers can be calculated from Eq. (2),

$$\begin{aligned} p(v|\theta) &= \sum_{h} p(v, h|\theta) = \frac{1}{Z}\sum_{h} e^{-E(v,h)} \\ p(h|\theta) &= \sum_{v} p(v, h|\theta) = \frac{1}{Z}\sum_{v} e^{-E(v,h)} \end{aligned} \tag{3}$$

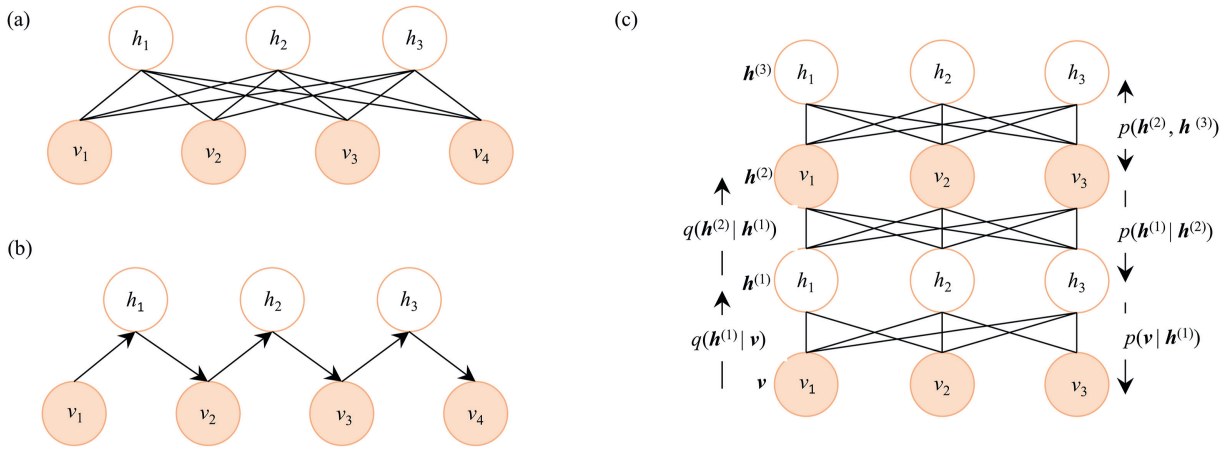

Fig. 1. The structure diagrams of RBM and DBN: (a) RBM, (b) RBM sampling process, (c) DBN.

The probability that the visible and hidden layer nodes are activated (*i.e.*, the neuron value is 1) is

$$\begin{aligned} p(h_j{=}\,1|v) &= \sigma\left(b_j + \sum_i^m w_{ij}v_i\right) \\ p(v_i{=}\,1|h) &= \sigma\left(a_i + \sum_j^n w_{ij}h_j\right) \end{aligned} \tag{4}$$

There are no connections between the nodes in the RBM layer and the neurons in the same layer are not related to each other, so there are

$$\begin{aligned} p(h|v) &= \prod_{j=1}^{n} p(h_j|v) \\ p(v|h) &= \prod_{i=1}^{m} p(v_i|h) \end{aligned} \tag{5}$$

2.2. Elastic weight consolidation

The training process of a neural network can be seen as optimizing the parameters θ over the training set $\mathscr{D}$ to maximize the conditional probability $p(\theta|\mathscr{D})$. According to the Bayesian criterion,

$$\lg p(\theta|\mathscr{D}) = \lg p(\mathscr{D}|\theta) + \lg p(\theta) - \lg p(\mathscr{D}). \tag{6}$$

The opposite of $\lg p(\mathscr{D}|\theta)$ in Eq. (6) is commonly known as the loss function. Dividing $\mathscr{D}$ into two independent subsets $\mathscr{D}_\mathrm{A}$ and $\mathscr{D}_\mathrm{B}$ yields

$$\lg p(\theta|\mathscr{D}_\mathrm{A}, \mathscr{D}_\mathrm{B}) = \lg p(\mathscr{D}_\mathrm{B}|\theta) + \lg p(\theta|\mathscr{D}_\mathrm{A}) - \lg p(\mathscr{D}_\mathrm{B}) \tag{7}$$

It is usually difficult to obtain an analytical solution for $\lg p(\theta|\mathscr{D}_\mathrm{A})$ or the analytical solution is very complex. In Ref. [37], the Laplace approximation is used, assuming that $\lg p(\theta|\mathscr{D}_\mathrm{A})$ follows a Gaussian distribution with mean θ_A^* and covariance matrix Σ satisfying $\Sigma^{-1} = \mathrm{diag}\{F_1, F_2, \cdots, F_d\}$. where d represents the number of parameters and $F_i = \left|(\partial L(\theta)/\partial\theta)|_{\theta_\mathrm{A}^*}\right|^2$ is the diagonal element of the Fisher information matrix of the parameter θ_A^*. The canonical form of the final EWC can be obtained by inserting F_i into Eq. (7).

$$L_\mathrm{EWC}(\theta) = L_\mathrm{B}(\theta) + \sum_i \frac{\lambda}{2} F_i \left(\theta_i - \theta_{\mathrm{A},i}^*\right)^2 \tag{8}$$

where λ is the hyperparameter and the superscript * represents the optimal value.

3. Proposed Method

3.1. Hard sample mining

Given a data set $\mathscr{D} = \{\boldsymbol{x}^{(t)}\}_{t=1}^{T}$, where T represents the number of sample points, the probability $p(h_j{=}\,1|\boldsymbol{x}^{(t)})$ can be calculated as

$$p\left(h_j{=}\,1|\boldsymbol{x}^{(t)}\right) = \sigma\left(b_j + \sum_i^m w_{ij}\boldsymbol{x}_i^{(t)}\right) \tag{9}$$

According to Eq. (5), sampling $\boldsymbol{h}$ yields the forward gradient as

$$\nabla^+ = \boldsymbol{x}^{(t)\mathrm{T}}\boldsymbol{h}^\mathrm{T} \tag{10}$$

The activation probability $p(v{=}\,1|\boldsymbol{h})$ is calculated as

$$p(v_i{=}\,1|\boldsymbol{h}) = \sigma\left(a_i + \sum_j^n w_{ij}h_j\right) \tag{11}$$

Collect the reconstructed $\boldsymbol{x}'^{(t)}$ according to $p(v{=}\,1|\boldsymbol{h})$, recompute the activation probability of the hidden layer neuron, sample the reconstructed $\boldsymbol{h}'$, and calculate the reverse gradient,

$$\nabla^- = \boldsymbol{x}'^{(t)\mathrm{T}}\boldsymbol{h}'^\mathrm{T} \tag{12}$$

Given the learning rate α, the parameters of a single RBM are updated in the following way,

$$\begin{aligned} \boldsymbol{W} &\leftarrow \boldsymbol{W} + \alpha(\nabla^+ - \nabla^-) \\ \boldsymbol{a} &\leftarrow \boldsymbol{a} + \alpha\left(\boldsymbol{x}^{(t)} - \boldsymbol{x}'^{(t)}\right) \\ \boldsymbol{b} &\leftarrow \boldsymbol{b} + \alpha(\boldsymbol{h} - \boldsymbol{h}') \end{aligned} \tag{13}$$

Assuming the DBN is constructed from stacked L-layer RBMs, the conditional probability of the hidden variable is computed as

$$p\left(\boldsymbol{h}^{(i)}\middle|\boldsymbol{h}^{(i-1)}\right) = \sigma\left(\boldsymbol{b}^{(i)} + \boldsymbol{W}^{(i)}\boldsymbol{h}^{(i-1)}\right), 1 \le i \le (L-1) \tag{14}$$

After training the DBN layer by layer following the procedure outlined in Eqs. (9)–(13), the parameters $\{\boldsymbol{W}^{(l)}, \boldsymbol{a}^{(l)}, \boldsymbol{b}^{(l)}\}, 1 \le l \le L$ are obtained. An output layer is added to the top layer of the DBN for fault diagnosis,

$$\boldsymbol{Z} = \boldsymbol{W}^{(L+1)}\boldsymbol{h}^{(L)} + \boldsymbol{b}^{(L+1)} \tag{15}$$

The prediction is

$$\widehat{Y}_i = \mathrm{Softmax}(\boldsymbol{z}_i) = \exp(\boldsymbol{z}_i)\left(\sum_{j=1}^{C}\exp(\boldsymbol{z}_j)\right)^{-1} \tag{16}$$

and the loss function is

$$L_{\mathrm{category}}(\boldsymbol{x},\boldsymbol{\theta}) = \sum_{i=1}^{C} Y_i \lg(\widehat{Y}_i) = \sum_{i=1}^{C} Y_i \lg(f(\boldsymbol{x},\boldsymbol{\theta})) \tag{17}$$

where $\boldsymbol{\theta} = \{\boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b}\}$ is the parameters and $f(\cdot)$ is the model to be trained. Next, the statistical distribution of sample loss is used for hard sample mining. Samples are categorized as hard $\mathscr{D}_{\mathrm{u}}$ or easy $\mathscr{D}_{\mathrm{e}}$ based on a defined threshold Q_β. Specifically,

$$\begin{aligned}\mathscr{D}_{\mathrm{e}} &= \left\{\boldsymbol{x}^{(t)} \middle| L\left(\boldsymbol{x}^{(t)},\boldsymbol{\theta}^*\right) \le Q_\beta\right\}\\ \mathscr{D}_{\mathrm{u}} &= \left\{\boldsymbol{x}^{(t)} \middle| L\left(\boldsymbol{x}^{(t)},\boldsymbol{\theta}^*\right) > Q_\beta\right\}\end{aligned} \tag{18}$$

The threshold Q_β is determined using kernel density estimation (KDE) of the sample loss distribution. The probability density function is estimated as:

$$F(x) = \frac{1}{n\pi}\sum_{i=1}^{n} K\left(\frac{x - x_i}{\pi}\right) \tag{19}$$

where $K(\cdot)$ is the Gaussian kernel function, π is the kernel width, n is the number of observed samples, and x_i is the individual sample loss. Based on the estimated density distribution, the loss value corresponding to the upper quantile γ ($\gamma = 0.9$) is selected as the threshold Q_β.

3.2. Domain adaptation

For D_{e}, the importance of each parameter in the model is calculated as follows:

$$\Omega = \frac{1}{|D_{\mathrm{e}}|}\sum_{\boldsymbol{x}\in D_{\mathrm{e}}}\frac{\partial^2 L(\boldsymbol{x},\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^{*2}} \tag{20}$$

However, calculating the importance of the above parameters is computationally expensive and can be approximated using the average of the squared gradients,

$$\Omega' = \frac{1}{|D_{\mathrm{e}}|}\sum_{\boldsymbol{x}\in D_{\mathrm{e}}}\frac{\partial L(\boldsymbol{x},\boldsymbol{\theta}^*)^2}{\partial \boldsymbol{\theta}^{*2}} \tag{21}$$

According to Eq. (21), the importance of all parameters is calculated and according to Eq. (18), the threshold Q_Ω is calculated. Parameters with importance greater than Q_Ω are restricted from changing as the model continues to be trained on D_{u}. Given that the occurrence of D_{u} may be due to feature drift in the feature space constructed by the model on such samples, it is essential to ensure that the feature for D_{u} remains consistent with that of D_{e}. Assuming that the output vectors of the hidden layer follow a Gaussian distribution, let the hidden layer vectors computed by f_θ for D_{e} and g_ϕ for D_{u} be $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively.

$$\begin{aligned}\boldsymbol{p} &= f_\theta(\boldsymbol{x},\theta) \sim \mathscr{N}\left(\boldsymbol{\mu}_p, \Sigma_p\right), \boldsymbol{x}\in\mathscr{D}_{\mathrm{e}}\\ \boldsymbol{q} &= g_\phi(\boldsymbol{x},\phi) \sim \mathscr{N}\left(\boldsymbol{\mu}_q, \Sigma_q\right), \boldsymbol{x}\in\mathscr{D}_{\mathrm{u}}\end{aligned} \tag{22}$$

where $\boldsymbol{\mu}_p$, Σ_p, $\boldsymbol{\mu}_q$ and Σ_q represent the mean and covariance matrix. We want to minimize the KL divergence between $\boldsymbol{p}$ and $\boldsymbol{q}$ by tuning the parameters of g_ϕ, ensuring that the features extracted by g_ϕ closely align with f_θ,

$$L_{\mathrm{DA}} = \min_{\phi} D_{\mathrm{KL}}\left(f_\theta(\boldsymbol{x}_{\mathrm{e}}), g_\phi(\boldsymbol{x}_{\mathrm{u}})\right), \boldsymbol{x}_{\mathrm{e}} \in D_{\mathrm{e}}, \boldsymbol{x}_{\mathrm{u}} \in D_{\mathrm{u}} \tag{23}$$

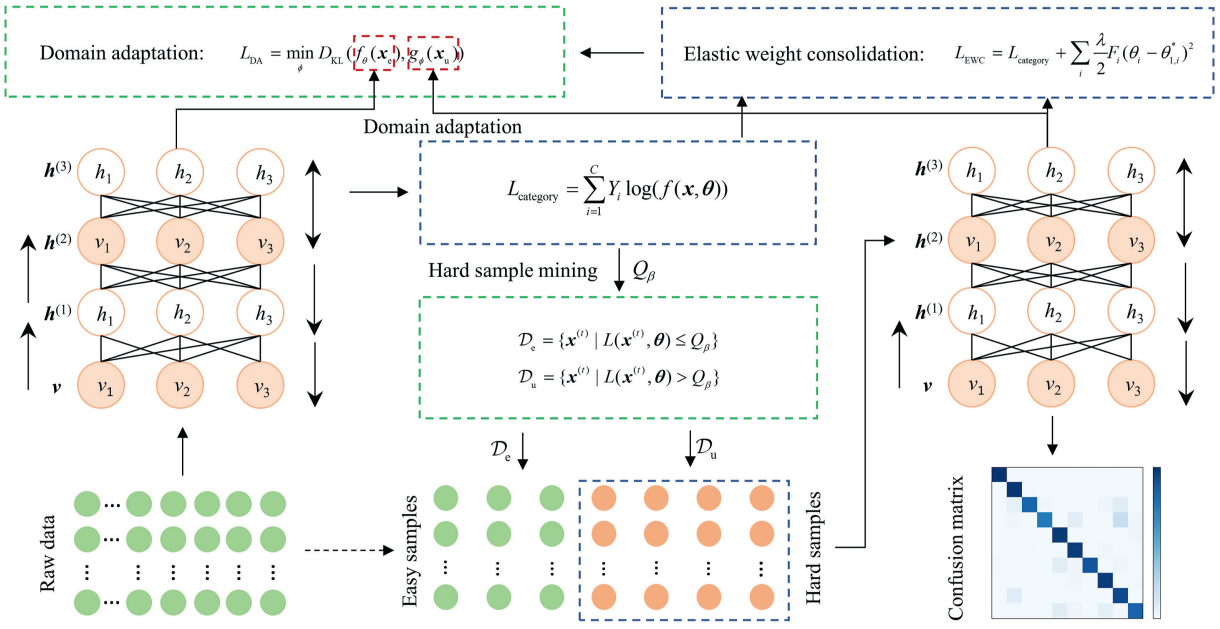

Fig. 2. Flowchart of SmdaNet.

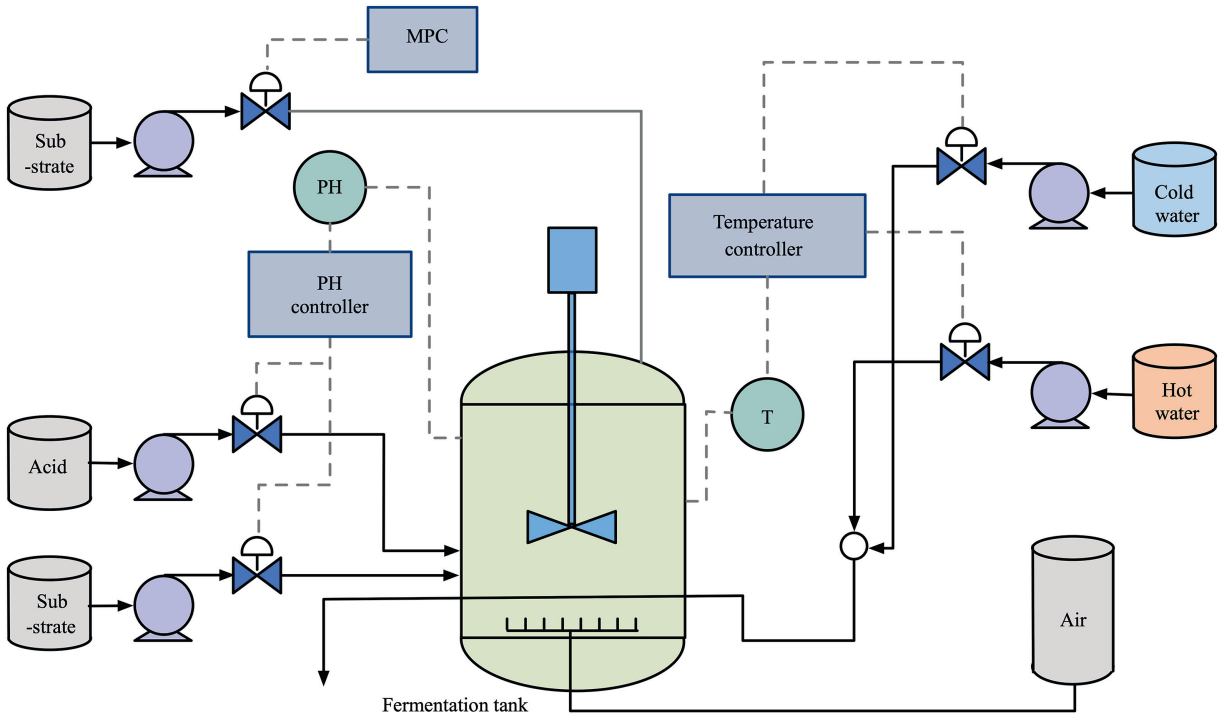

Fig. 3. Diagram of penicillin simulation process.

Table 1
The selected ten variables for analysis.

No.	Variable description	No.	Variable description
x_1	Aeration rate	x_6	DO
x_2	Agitator power	x_7	Culture volume
x_3	Substrate feed flow rate	x_8	CO_2
x_4	Substrate feed temperature	x_9	pH
x_5	Substrate concentration	x_{10}	Temperature

Table 2
Ten faults used in the penicillin case.

No.	Fault variable	Fault type	Fault size
1	Agitator power	Step change	+4%
2	Agitator power	Step change	−2%
3	Agitator power	Ramp change	+2 W
4	Agitator power	Ramp change	−4 W
5	Aeration rate	Step change	+2%
6	Aeration rate	Step change	−3%
7	Aeration rate	Ramp change	+2 L·h^{-1}
8	Aeration rate	Ramp change	−1 L·h^{-1}
9	Substrate feed flow rate	Step change	+8%
10	Substrate feed flow rate	Step change	−8%

Table 3
Values of main parameters of penicillin simulation process modeling.

	MLP	VAE	DBN	EDBN	SurvNet	SmdaNet
lr	0.01	0.01	0.02	0.05	0.01	0.02
B	32	32	64	64	16	32
L	1	2	2	2	2	2
d	0.1	0.1	0.1	0.1	0.1	0.1
η	–	–	–	–	0.2	–

Eq. (23) facilitates domain adaptation to mitigate the effects of potential drift and improve the generalization ability of the model. The total loss function on the D_u is then expressed as

Table 4
Hyperparameter π sensitivity.

	0.1	0.15	0.2	0.25	0.3
MF	0.9866	0.9865	0.9870	0.9866	0.9867

Table 5
The macro average precision, recall, F1-score and accuracy results.

	MP	MR	MF	MA
MLP	0.7972	0.7090	0.6903	0.7090
VAE	0.8426	0.8190	0.8153	0.8190
DBN	0.8920	0.8870	0.8865	0.8870
SurvNet	0.9040	0.8980	0.8977	0.8980
EDBN	0.9257	0.9240	0.9243	0.9240
SmdaNet	**0.9870**	**0.9870**	**0.9869**	**0.9870**

$$L_{\text{total}} = L_{\text{DA}} + L_{\text{category}} + L_{\text{EWC}} \tag{24}$$

After training $f_\theta(\cdot)$ and $g_\phi(\cdot)$ according to L_{DA} and L_{total}, they can be used to perform the multi-classification task on the test set. The flowchart of SmdaNet is shown in Fig. 2.

4. Case Studies

4.1. Penicillin simulation process

The simulation data used in this study were obtained from the penicillin production simulation platform developed by Birol *et al.* [38], as shown in Fig. 3.

To validate the effectiveness of SmdaNet in identifying faults within the penicillin production process, we selected 10 variables (Table 1) and introduced 10 faults (Table 2) for subsequent analysis.

Macro-mean precision (MP), macro-mean recall (MR), macro-mean F1-score (MF), and macro-mean accuracy (MA) are chosen as the evaluation metrics of the model.

$$\begin{aligned} \mathrm{MP} &= \frac{1}{C}\sum_{i=1}^{C} P_i, \qquad P_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i} \\ \mathrm{MR} &= \frac{1}{C}\sum_{i=1}^{C} R_i, \qquad R_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i} \\ \mathrm{MF} &= \frac{1}{C}\sum_{i=1}^{C} \mathrm{F1}_i, \qquad \mathrm{F1}_i = \frac{2\times \mathrm{TP}_i}{2\times \mathrm{TP}_i + \mathrm{FP} + \mathrm{FN}_i} \\ \mathrm{MA} &= \frac{1}{C}\sum_{1}^{C} \mathrm{Acc}_i, \quad \mathrm{Acc}_i = \frac{\mathrm{TP}_i + \mathrm{TN}_i}{\mathrm{TP}_i + \mathrm{TN}_i + \mathrm{FP}_i + \mathrm{FN}_i} \end{aligned} \tag{25}$$

where C is the number of categories, and subscript "i" represents the category. For each category, 1000 samples are generated to construct the training and test sets, respectively. The main parameters of the model are determined using grid search, as detailed in Table 3.

lr, B, L, d represent the learning rate, batch size, number of network layers, and dropout value, respectively; η represents the false discovery rate, which is used to determine the termination of SurvNet [39].

To ensure the robustness of threshold determination, we conduct a sensitivity analysis on the kernel width π using grid search within a reasonable range of [0.1,0.3], as shown in Table 4. The results show that the model performance exhibits only slight fluctuations across this interval, indicating low sensitivity to this parameter. Therefore, π is fixed at 0.2 in all subsequent experiments for a balance between stability and empirical performance.

The average classification results of the model, obtained from 20 simulated experiments, are presented in Table 5, with detailed results illustrated in Fig. 4.

Table 5 summarizes the average results of 20 experiments conducted on six models across four key performance indicators. The traditional MLP and VAE methods show significantly lower performance compared to other approaches. MLP scores the lowest across all metrics, highlighting its limited ability to extract features. VAE, which uses latent space representations, shows slight improvements in MF and MA metrics, but falls short of satisfactory performance. DBN and SurvNet use deep learning to

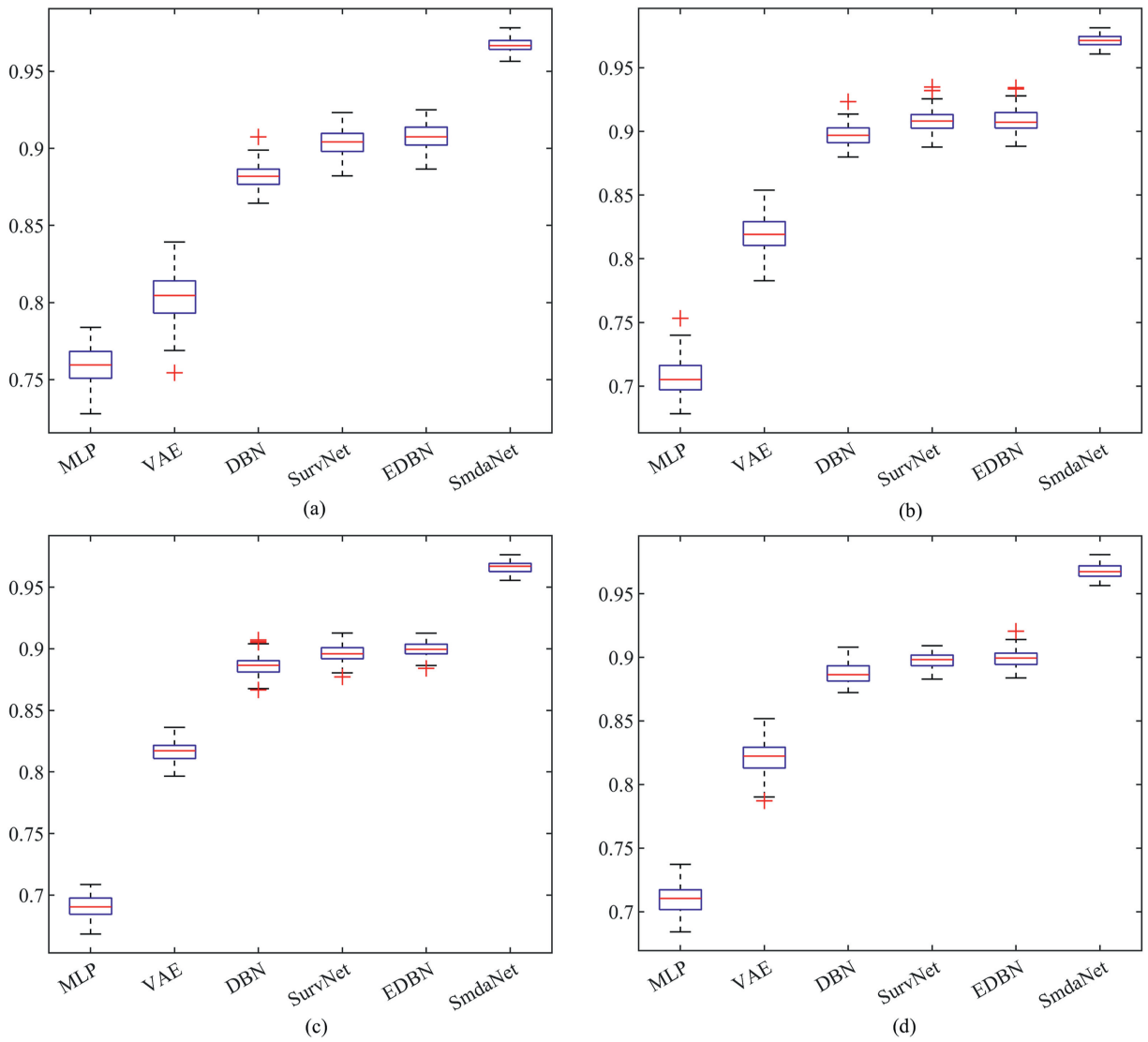

Fig. 4. Box plots of classification results: (a) MP, (b) MR, (c) MF, (d) MA.

Table 6
Comparison of model computational efficiency (s).

Sample number	DBN	EDBN	Δ_t	SmdaNet/s	Δ_t
200	215	312	97	226	21
400	246	361	115	255	19
600	264	426	162	273	19
800	306	501	195	322	26
1000	315	633	318	364	49

Table 7
Ablation experiments for EWC and DA.

	SmdaNet	EWC		DA	
		w/o	Δ	w/o	Δ
MP	0.9870	0.9034	−9.25%	0.9204	−7.24%
MR	0.9870	0.8930	−10.53%	0.9090	−8.58%
MF	0.9869	0.8927	−10.55%	0.9108	−8.36%
MA	0.9870	0.8930	−10.53%	0.9090	−8.58%

effectively extract non-linear features, achieving significant improvements over traditional models and highlighting the benefits of hierarchical feature extraction. EDBN improves MF and MA to 0.9243 and 0.9240 respectively by incorporating raw data into the network structure. This highlights the critical role of raw data in improving feature representation and model robustness. SmdaNet outperforms EDBN, achieving MF and MA scores of 0.9870 and 0.9869 respectively, demonstrating its superior modelling capabilities.

The box plots in Fig. 4 show that MLP and VAE have wider distributions and more outliers, highlighting their poor robustness. In contrast, DBN, SurvNet and EDBN show comparable performance but fall short of SmdaNet. SmdaNet has the most concentrated distributions across all metrics, with minimal outliers and exceptionally short boxes in MR and MF, indicating high stability. In addition, SmdaNet achieves the highest medians across all four metrics, highlighting its consistent performance, superior accuracy and robustness in experiments.

Table 6 compares the computational efficiency of DBN, EDBN and SmdaNet for different sample sizes, highlighting the additional computational time compared to DBN. While SmdaNet has a slightly higher computational time than DBN, it significantly outperforms EDBN in terms of efficiency, especially for large sample sizes (*e.g.* 1000 samples). For example, SmdaNet requires 364 s - only 49 s more than DBN - while EDBN requires an additional 318 s. These results highlight SmdaNet's ability to improve performance while maintaining computational efficiency, making it more resource-efficient and suitable for engineering applications than EDBN.

The ablation experiments in Table 7 highlight the critical contributions of the EWC and DA modules to the performance of SmdaNet. Removing the EWC module results in a significant decrease in all metrics, exceeding 10%, highlighting its role in preserving existing knowledge and preventing information forgetting. Similarly, removing the DA module leads to a decrease in metrics ranging from 7.24% to 8.58%, highlighting its importance in feature optimization and category differentiation. A comparative analysis shows that the EWC module primarily improves model stability and robustness, while the DA module focuses on improving feature extraction and classification performance. The synergy of these modules enables SmdaNet to achieve superior classification accuracy and stability, confirming the effectiveness of the module design.

Fig. 5 illustrates the MF metric trends and fluctuations with increasing sample size, showing the performance of the methods under both small and large sample conditions. In small sample scenarios, MLP and VAE show lower MF values and significant

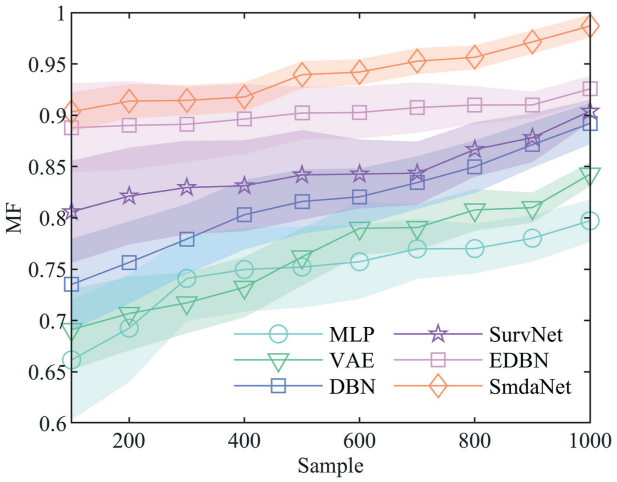

Fig. 5. Variations of MF metrics with sample size.

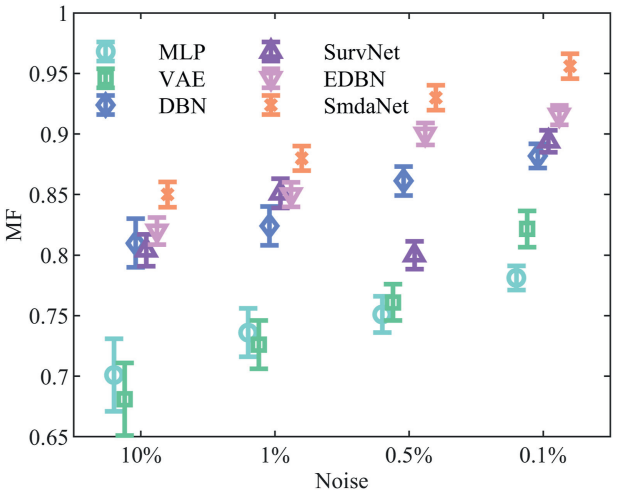

Fig. 6. Variations of model effect with noise intensity.

fluctuations, reflecting poor stability. DBN shows improved stability with reduced variability, while SurvNet, EDBN and SmdaNet perform significantly better. In particular, SmdaNet achieves an initial MF value of 0.91, which rises steadily to 0.98 as the sample size increases, with minimal variation. This highlights the exceptional stability and robustness of SmdaNet across different sample sizes, making it highly adaptable to different tasks.

Fig. 6 illustrates the MF performance and error ranges of the different methods under different noise levels (10%, 1%, 0.5%, 0.1%). MLP and VAE show poor performance under high noise, with MF values around 0.70 and 0.68 at 10% noise, respectively. Their long error bars indicate significant performance fluctuations and weak robustness. In contrast, DBN, SurvNet and EDBN show improved performance and robustness. EDBN consistently outperforms DBN at all noise levels, showing stronger resistance to noise. SmdaNet consistently achieves the best performance, with an MF of around 0.85 at 10% noise, improving to 0.95 at 0.1% noise. It has the smallest error range, indicating exceptional stability and robustness. This highlights the ability of SmdaNet to effectively mitigate noise interference while maintaining efficient feature extraction capabilities. Overall, SmdaNet outperforms other methods in terms of noise robustness and result stability, making it well suited for practical applications in complex noise environments.

Fig. 7 shows the confusion matrices of six models, comparing their classification performance on 100 selected samples. Darker squares along the diagonal represent correctly classified samples, while lighter squares off-diagonal represent misclassified samples. As model complexity increases (Fig. 7(a)–(f)), the diagonal values increase and the off-diagonal values decrease, reflecting a steady improvement in classification accuracy. MLP and VAE show lower classification performance, with significant misclassification problems, especially in category 5. In contrast, DBN, SurvNet and EDBN show significant improvements in classification accuracy in most categories, although misclassifications persist in categories 5, 8 and 9. SmdaNet achieves the best performance, with the highest classification accuracy in almost all categories. It reduces the number of misclassified samples to the lowest level, demonstrating exceptional performance and reliability. These results highlight the significant advantages of SmdaNet in handling complex classification tasks, making it highly suitable for such challenges.

Fig. 8 shows the visualization results of SmdaNet for different categories, with each color representing a different category of data points. SmdaNet achieves completely separated category

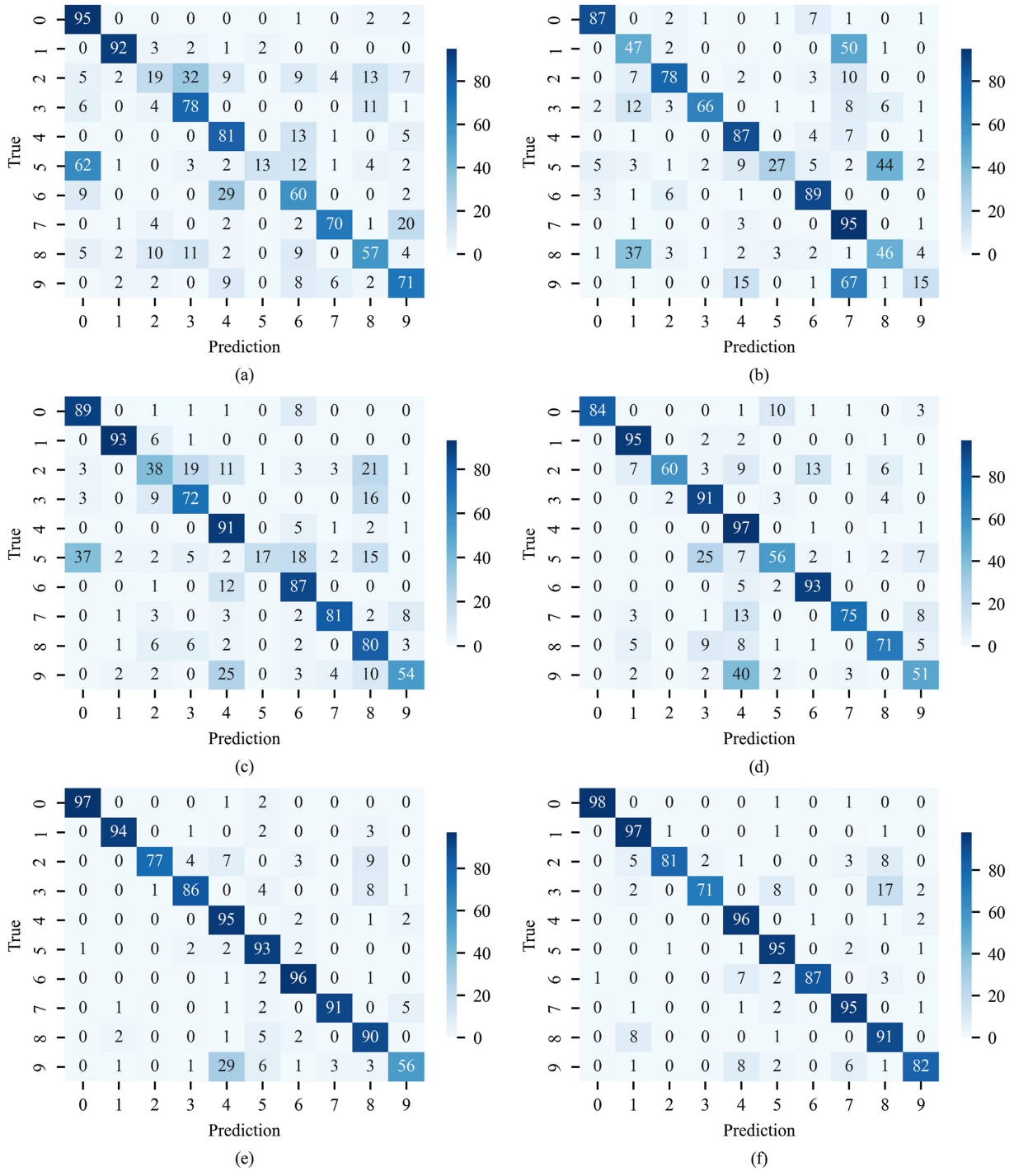

Fig. 7. Confusion matrixes for penicillin simulation data: (a) MLP, (b) VAE, (c) DBN, (d) SurvNet, (e) EDBN, (f) SmdaNet.

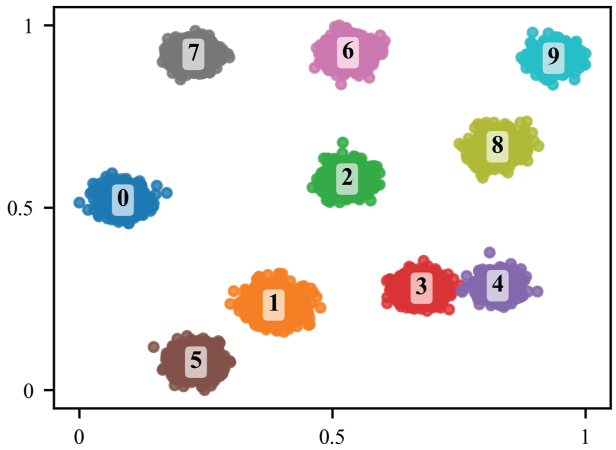

Fig. 8. Visualization results of SmdaNet for penicillin simulation dataset.

distributions with well-defined boundaries and no significant feature overlap. The visualization results closely match the confusion matrix analysis in Fig. 7, highlighting SmdaNet's superior capability in feature extraction and category differentiation. It effectively captures complex data features, generates highly separated feature spaces, and excels in dimensionality reduction and category differentiation for high-dimensional data.

4.2. TE

Fig. 9 illustrates the TE process, a chemical modelling simulation platform developed by Eastman Chemical Company (USA) based on real chemical reaction processes [40]. The TE simulation platform generates data with time-varying, highly coupled and non-linear characteristics.

Table 8 details the variables and fault information used in the simulation and Table 9 shows the sample sizes for training and testing.

Table 8
Variables used in the TE process.

	No.	Variable	Description
Variable	1–22	XMEAS	Measurement
	23–33	XMV	Manipulated
Fault	0	–	Normal
	1–21	IDV	Fault

Table 9
Number of samples used in the TE process.

	Training	Test
Normal	500	960
Fault(1–21)	480	160 (normal)+800 (fault)

Table 10
Values of the main model parameters of the TE process.

	MLP	VAE	DBN	EDBN	SurvNet	SmdaNet
lr	0.005	0.005	0.01	0.01	0.01	0.01
B	16	16	32	32	16	16
L	2	2	3	3	2	2
d	0.1	0.1	0.1	0.1	0.1	0.1
η	–	–	–	–	0.2	–

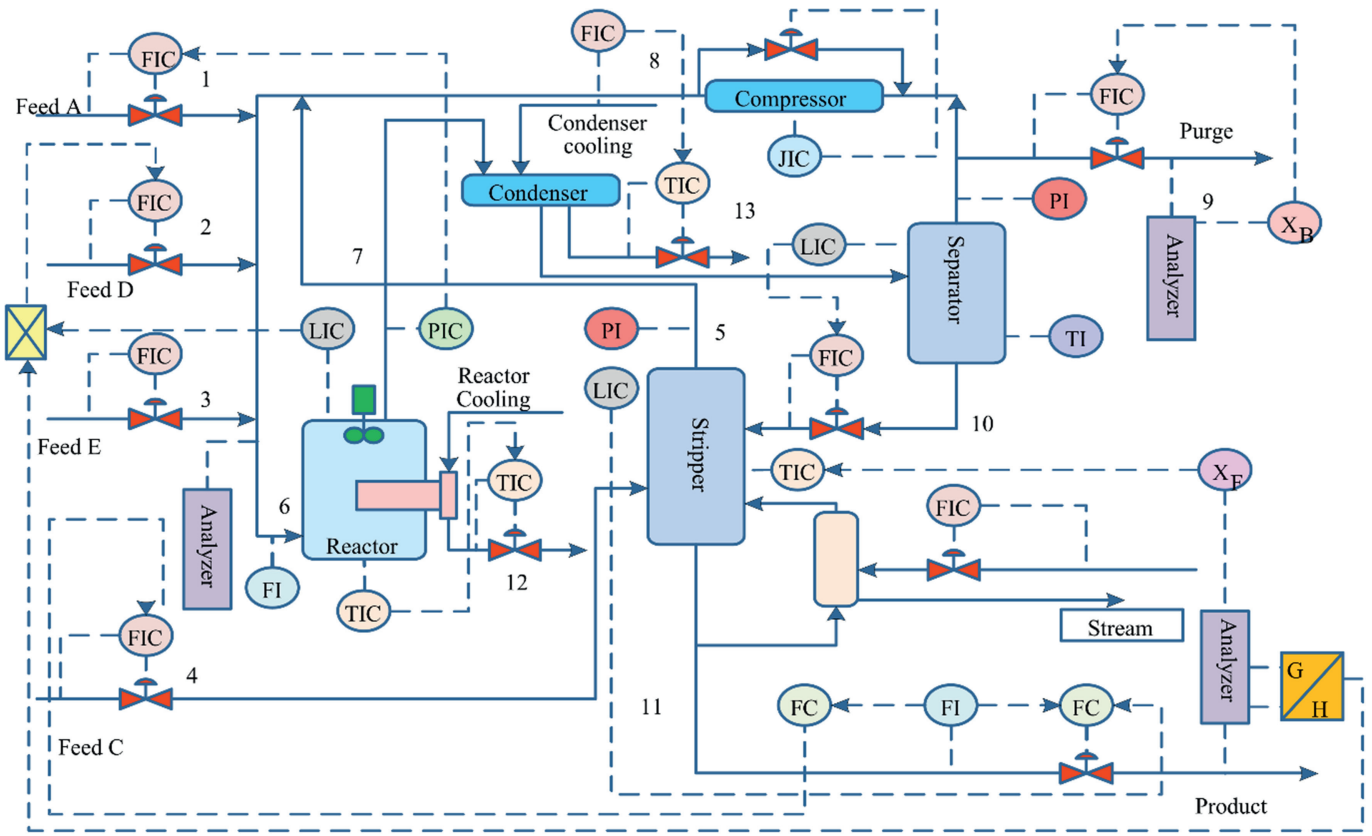

Fig. 9. The diagram of Tennessee Eastman process.

Table 11
Comparison of model effectiveness (precision) of TE process.

Fault type	MLP	VAE	DBN	EDBN	SurvNet	SmdaNet
0	0.37	0.55	0.75	0.82	**0.9**	0.79
1	**1**	**1**	0.99	0.73	0.94	**1**
2	0.74	0.99	**1**	0.99	0.99	**1**
3	0.94	0.97	**1**	**1**	**1**	0.68
4	0.51	0.73	**1**	0.84	0.99	**1**
5	**1**	0.45	0.61	**1**	**1**	**1**
6	0.86	0.94	0.97	0.98	0.99	**1**
7	0.82	0.43	**1**	0.6	0.96	**1**
8	**1**	0.45	0.97	0.91	0.59	**1**
9	0.97	**1**	0.97	**1**	0.94	0.68
10	0.99	0.72	0.84	0.85	**1**	0.98
11	**1**	**1**	**1**	0.99	**1**	0.96
12	**1**	0.71	**1**	**1**	**1**	**1**
13	0.95	0.94	0.95	0.82	0.94	**1**
14	0.99	**1**	0.99	0.99	**1**	**1**
15	0.61	**1**	0.65	**1**	0.8	0.79
16	0.63	0.98	0.99	**1**	**1**	1
17	0.46	**1**	0.8	0.87	**1**	1
18	0.52	0.91	0.66	**1**	0.79	0.98
19	**1**	**1**	0.99	0.97	0.95	0.96
20	0.9	**1**	**1**	**1**	**1**	0.99
21	**1**	0.95	0.96	**1**	0.98	0.99
Average	0.83	0.85	0.91	0.93	0.94	0.95
Best number	7	8	7	9	9	10

Table 10 shows the parameters used in the TE process. Table 11 summarizes the classification performance of six models across 21 fault types and one normal condition, along with their average scores and the number of categories in which each model performed best (best number). MLP achieves the lowest average accuracy of all models, due to limited feature extraction capacity from its shallow architecture. VAE performs well on a few fault types but suffers from significant drops in accuracy (*e.g.*, types 5, 7, and 8). DBN, EDBN and SurvNet show a good overall performance and scores well for several fault types, but still performs poorly in diagnosing some fault types. In terms of average performance, SmdaNet outperforms the other five models with a score of 0.95. In addition, SmdaNet shows remarkable robustness, with minimal performance variability and consistently high classification accuracy (close to 1) across most types. For the best number of classifications, SmdaNet achieves optimal performance in 10 categories, compared to 9 for SurvNet and 7 for EDBN. In contrast, DBN, MLP and VAE show comparatively lower performance. These results highlight SmdaNet's exceptional ability to handle category diversity and task complexity, making it a comprehensive and efficient fault diagnosis method for addressing complex industrial requirements.

Fig. 10 shows the confusion matrix of the SmdaNet for the TE process, showing minor misclassifications in category 15.

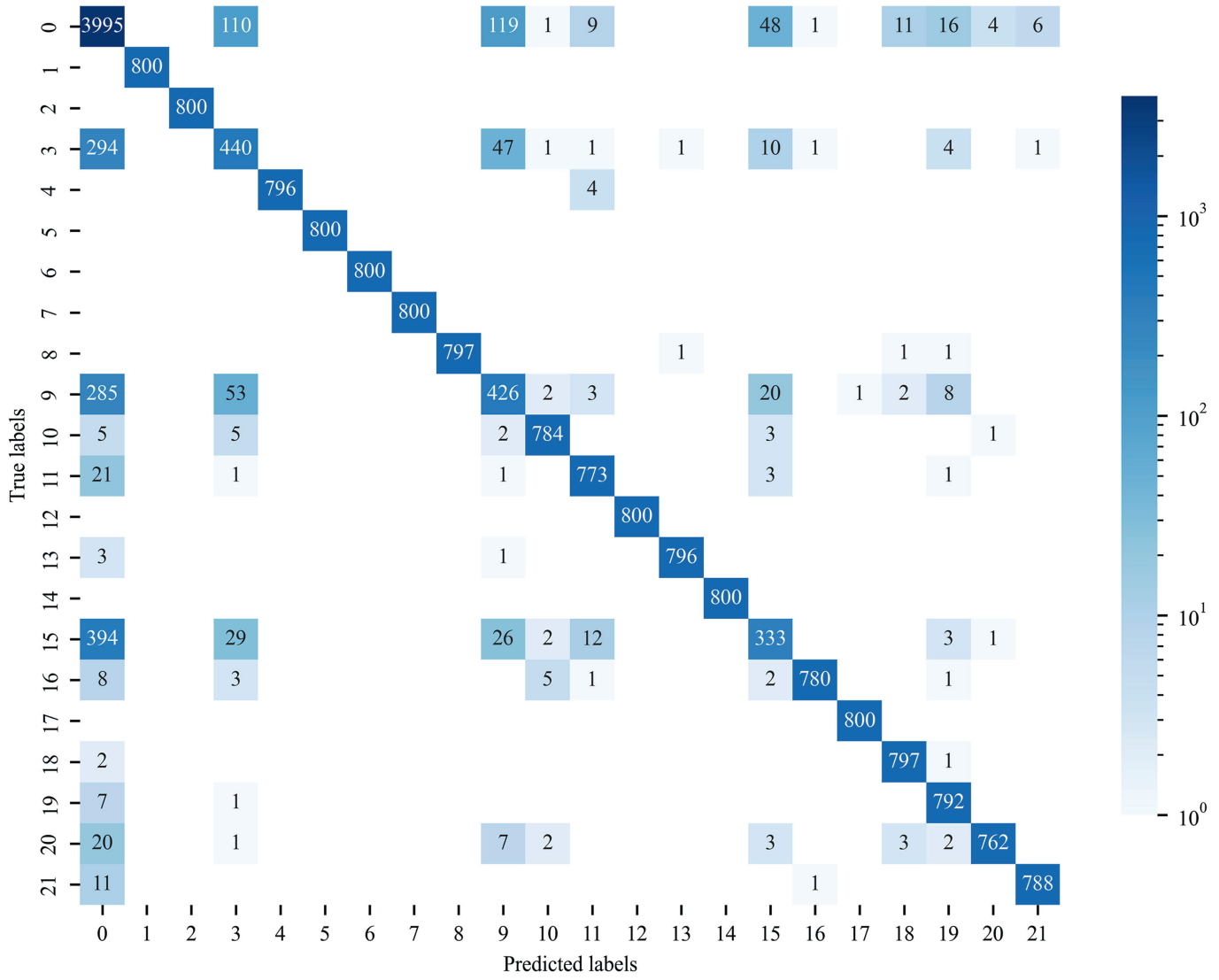

Fig. 10. Confusion matrix of SmdaNet on TE process.

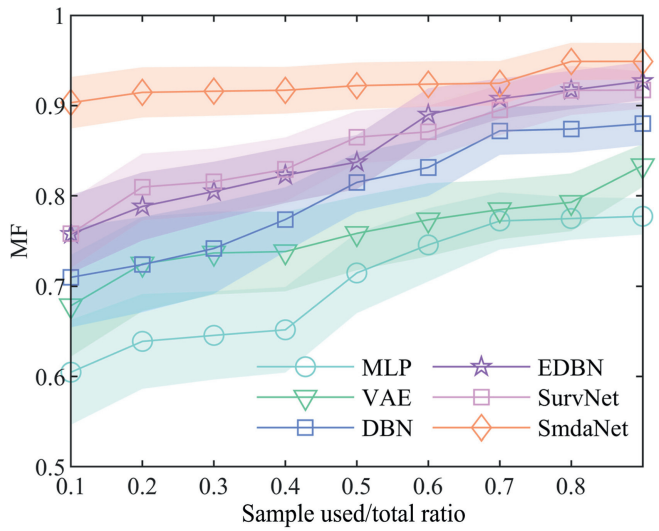

Fig. 11. The effects of 6 models varies with the number of samples in the TE process.

Nevertheless, the most of the samples are concentrated along the diagonal, indicating a high overall classification performance. In particular, the model achieves near perfect accuracy for labels 5, 6, 7 and 8. Overall, SmdaNet shows superior classification performance in most categories, with minimal areas for improvement.

Fig. 11 shows the evolution of the MF values for the models on TE process data as the percentage of samples increases. While all methods show significant improvements in MF values with increasing sample fractions, the differences in performance become more pronounced. MLP and VAE consistently underperform, particularly at low sample fractions, with MF values around 0.6 and 0.68. In contrast, DBN, SurvNet and EDBN show significant performance gains. SmdaNet outperforms all methods, achieving MF values close to 0.9 at low sample fractions and 0.95 at high sample fractions, with minimal variation, highlighting its superior stability and robustness. These results underline the significant advantages of SmdaNet in terms of classification performance and stability compared to the other five methods.

By combining experiments on penicillin simulation data and TE process data, SmdaNet demonstrates optimal classification performance, stability and robustness across both datasets. In terms of confusion matrices and classification results, SmdaNet consistently outperforms other methods in accuracy across all categories, with minimal misclassification. Noise experiments further validate its robustness, as SmdaNet maintains high classification performance under noisy conditions and converges quickly to the optimum as the noise decreases. In addition, SmdaNet excels in low-sample-ratio scenarios, making efficient use of limited data. Its computational overhead remains low relative to its performance gains, significantly reducing resource consumption. Feature visualization results show that SmdaNet achieves clear category separation with well-defined classification boundaries, highlighting its robust feature extraction and nonlinear representation capabilities.

5. Conclusions

This paper presents SmdaNet, a fault diagnosis method based on DBN designed to overcome the limitations of handling hard samples and ensuring cross-domain feature consistency. The SmdaNet categorizes training data into hard and easy samples based on loss values. EWC is used to fine-tune the model on hard samples, improving its classification accuracy while preserving knowledge from easy samples. In addition, a feature space adaptation mechanism minimizes the KL divergence between the feature distributions of hard and easy samples, improving feature representation and cross-domain adaptability. The experiments on penicillin simulation data and TE process data demonstrate the superiority of SmdaNet. Although SmdaNet showed superior performance, it has two major limitations. First, it relies on labeled source domain data, which can be costly and impractical to obtain in early industrial deployments. Second, its performance may degrade under significant domain shifts where the discrepancy between the source and target distributions is large. To address these issues, future research will explore semi-supervised learning to reduce label dependence, adversarial strategies to improve domain robustness.

CRediT Authorship Contribution Statement

Zhenhua Yu: Writing – original draft, Validation, Methodology, Investigation. Zongyu Yao: Validation, Resources, Methodology. Weijun Wang: Validation, Resources, Methodology. Qingchao Jiang: Writing – review & editing, Supervision, Methodology. Zhixing Cao: Writing – review & editing, Supervision, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the support from the following foundations: the National Natural Science Foundation of China (62322309, 62433004), Shanghai Science and Technology Innovation Action Plan (23S41900500), and Shanghai Pilot Program for Basic Research (22TQ1400100-16).

References

[1] S.W. Xiong, L. Zhou, Y.Y. Dai, X. Ji, Attention-based long short-term memory fully convolutional network for chemical process fault diagnosis, *Chin. J. Chem. Eng.* 56 (2023) 1–14.
[2] J.J. Luo, Z.H. Jin, H.P. Jin, Q. Li, X. Ji, Y.Y. Dai, Causal temporal graph attention network for fault diagnosis of chemical processes, *Chin. J. Chem. Eng.* 70 (2024) 20–32.
[3] C.T. Wang, H.B. Shi, B. Song, Y. Tao, Hierarchical multihead self-attention for time-series-based fault diagnosis, *Chin. J. Chem. Eng.* 70 (2024) 104–117.
[4] C.B. Low, D.W. Wang, S. Arogeti, J.B. Zhang, Causality assignment and model approximation for hybrid bond graph: fault diagnosis perspectives, *IEEE Trans. Autom. Sci. Eng.* 7 (3) (2010) 570–580.
[5] L.U. Iurinic, A.R. Herrera-Orozco, R.G. Ferraz, A.S. Bretas, Distribution systems high-impedance fault location: a parameter estimation approach, *IEEE Trans. Power Deliv.* 31 (4) (2016) 1806–1814.
[6] A. Izadian, P. Khayyer, Application of Kalman filters in model-based fault diagnosis of a DC-DC boost converter, in: IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, IEEE, Glendale, AZ, USA, 2010.
[7] A. Ben Youssef, S.K. El Khil, I. Slama-Belkhodja, State observer-based sensor fault detection and isolation, and fault tolerant control of a single-phase PWM rectifier for electric railway traction, *IEEE Trans. Power Electron.* 28 (12) (2013) 5842–5853.
[8] S. Byun, M. Papaelias, F.P.G. Márquez, D. Lee, Fault-tree-analysis-based health monitoring for autonomous underwater vehicle, *J. Mar. Sci. Eng.* 10 (12) (2022) 1855.
[9] Y. Wilhelm, P. Reimann, W. Gauchel, B. Mitschang, Overview on hybrid approaches to fault detection and diagnosis: combining data-driven, physics-based and knowledge-based models, *Proced. CIRP* 99 (2021) 278–283.
[10] E.M. Kelly, L.M. Bartlett, Application of the digraph method in system fault diagnostics, First International Conference on Availability, Reliability and Security, Vienna, Austria, 2006.
[11] M. Misra, H.H. Yue, S.J. Qin, C. Ling, Multivariate process monitoring and fault diagnosis by multi-scale PCA, *Comput. Chem. Eng.* 26 (9) (2002) 1281–1293.
[12] G. Li, S.Z. Qin, Y.D. Ji, D.H. Zhou, Total PLS based contribution plots for fault diagnosis, *Acta Autom. Sin.* 35 (6) (2009) 759–765.

[13] Q.C. Jiang, X.F. Yan, J. Li, PCA-ICA integrated with Bayesian method for Non-Gaussian fault diagnosis, *Ind. Eng. Chem. Res.* 55 (17) (2016) 4979–4986.
[14] S.W. Choi, C. Lee, J.M. Lee, J.H. Park, I.B. Lee, Fault detection and identification of nonlinear processes based on kernel PCA, *Chemometr. Intell. Lab. Syst.* 75 (1) (2005) 55–67.
[15] J.M. Lee, S. Joe Qin, I.B. Lee, Fault detection of non-linear processes using kernel independent component analysis, *Can. J. Chem. Eng.* 85 (4) (2007) 526–536.
[16] S. Yin, X.P. Zhu, O. Kaynak, Improved PLS focused on key-performance-indicator-related fault diagnosis, *IEEE Trans. Ind. Electron.* 62 (3) (2015) 1651–1658.
[17] K. Zhong, M. Han, T. Qiu, B. Han, Fault diagnosis of complex processes using sparse kernel local fisher discriminant analysis, *IEEE Transact. Neural Networks Learn. Syst.* 31 (5) (2020) 1581–1591.
[18] A. Hajnayeb, A. Ghasemloonia, S.E. Khadem, M.H. Moradi, Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis, *Expert Syst. Appl.* 38 (8) (2011) 10205–10209.
[19] R.V. Sanchez, P. Lucero, R.E. Vasquez, M. Cerrada, J.C. Macancela, D. Cabrera, Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN, *J. Intell. Fuzzy Syst.* 34 (6) (2018) 3463–3473.
[20] N. Saravanan, V.N.S.K. Siddabattuni, K.I. Ramachandran, Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM), *Appl. Soft Comput.* 10 (1) (2010) 344–360.
[21] Y.T. Dong, H.K. Jiang, Z.H. Wu, Q. Yang, Y.P. Liu, Digital twin-assisted multi-scale residual-self-attention feature fusion network for hypersonic flight vehicle fault diagnosis, *Reliab. Eng. Syst. Saf.* 235 (2023) 109253.
[22] M.Z. Mu, H.K. Jiang, X. Wang, Y.T. Dong, A task-oriented theil index-based meta-learning network with gradient calibration strategy for rotating machinery fault diagnosis with limited samples, *Adv. Eng. Inform.* 62 (2024) 102870.
[23] X. Wang, H.K. Jiang, M.Z. Mu, Y.T. Dong, A dynamic collaborative adversarial domain adaptation network for unsupervised rotating machinery fault diagnosis, *Reliab. Eng. Syst. Saf.* 255 (2025) 110662.
[24] S.S. Zhong, S. Fu, L. Lin, A novel gas turbine fault diagnosis method based on transfer learning with CNN, *Measurement* 137 (2019) 435–453.
[25] B. Zhao, X.M. Zhang, H. Li, Z.B. Yang, Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions, *Knowl. Base Syst.* 199 (2020) 105971.
[26] W. Zhang, G.L. Peng, C.H. Li, Y.H. Chen, Z.J. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors* 17 (2) (2017) 425.
[27] H.T. Zhao, S.Y. Sun, B. Jin, Sequential fault diagnosis based on LSTM neural network, *IEEE Access* 6 (2018) 12929–12939.
[28] C. Zhang, Q.X. Zhu, Y.L. He, Y. Zhang, M.Q. Zhang, Y. Xu, Deep graph convolutional neural network for fault diagnosis of complex industrial processes, in: 2023 2nd International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics (AIHCIR), IEEE, Tianjin, China, 2023.
[29] K. Zhou, E. Diehl, J. Tang, Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations, *Mech. Syst. Signal Process.* 185 (2023) 109772.
[30] G. San Martin, E.L. Droguett, Temporal variational auto-encoders for semi-supervised remaining useful life and fault diagnosis, *IEEE Access* 10 (2022) 55112–55125.
[31] Z.H. Wang, T. Sun, X.C. Tian, Fault diagnosis of rolling bearing based on SDAE and PSO-DBN, in: 2019 Chinese Control and Decision Conference (CCDC), IEEE, Nanchang, China, 2019.
[32] W. Deng, H.L. Liu, J.J. Xu, H.M. Zhao, Y.J. Song, An improved quantum-inspired differential evolution algorithm for deep belief network, *IEEE Trans. Instrum. Meas.* 69 (10) (2020) 7319–7327.
[33] Z.Y. Chen, W.H. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, *IEEE Trans. Instrum. Meas.* 66 (7) (2017) 1693–1702.
[34] H. Oh, J.H. Jung, B.C. Jeon, B.D. Youn, Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis, *IEEE Trans. Ind. Electron.* 65 (4) (2018) 3539–3549.
[35] J.H. Yan, Y.Y. Hu, C.Z. Guo, Rotor unbalance fault diagnosis using DBN based on multi-source heterogeneous information fusion, *Procedia Manuf.* 35 (2019) 1184–1189.
[36] G.X. Niu, X. Wang, M. Golda, S. Mastro, B. Zhang, An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis, *Neurocomputing* 445 (2021) 26–34.
[37] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A.A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, R. Hadsell, Overcoming catastrophic forgetting in neural networks, *Proc. Natl. Acad. Sci. USA* 114 (13) (2017) 3521–3526.
[38] G. Birol, Ü. Cenk, Ç. Ali, A modular simulation package for fed-batch fermentation: penicillin production, *Comput. Chem. Eng.* 26 (11) (2002) 1553–1565.
[39] Z.X. Song, J. Li, Variable selection with false discovery rate control in deep neural networks, *Nat. Mach. Intell.* 3 (2021) 426–433.
[40] Y.L. Wang, Z.F. Pan, X.F. Yuan, C.H. Yang, W.H. Gui, A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network, *ISA Trans.* 96 (2020) 457–467.