Chinese Journal of Chemical Engineering 90 (2026) 181–195

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Sample-optimized adaptive perceptual enhanced graph neural network based fault identification in industrial processes

Keyu Yao , Hongbo Shi *, Bing Song , Yang Tao

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 August 2024
Received in revised form
8 September 2025
Accepted 20 October 2025
Available online 31 October 2025

Keywords:
Adaptive perception
Deep learning
Enhanced attention mechanism
Fault diagnosis
Sample-optimized feature processing strategy
Chemical process

ABSTRACT

With the complexity and intelligence of the industrial process, the identification of faults in the actual process plays a crucial role in ensuring production safety. The traditional fault identification strategies have the problem that similar characterized faults are unable to be accurately identified. Motivated by the limitations, a novel sample-optimized adaptive perceptual enhanced graph neural network (SOAP-EGNN) for large-scale process fault identification is proposed. Initially, process mechanism knowledge and process data correlation are injected into the modeling approach through graph neural networks, and the transmission of information based on the enhanced attention mechanism is introduced to describe the quantitative relationships between process variables at a fine-grained level based on the adaptive perception strategy. Subsequently, to achieve better intra-class compactness and inter-class separability in feature representation, our designed sample-optimized feature processing strategy (SOFPS) is applied. Furthermore, to enhance the robustness and generalization capability of the model during training, a label smoothing regularization (LSR) strategy is incorporated. This approach effectively mitigates the risk of overfitting by introducing a degree of uncertainty into the label space, thereby encouraging the model to learn more discriminative and stable features. Ultimately, the efficacy and superiority of the SOAP-EGNN algorithm are thoroughly validated through comprehensive simulation experiments conducted on the Tennessee Eastman process (TEP).

1. Introduction

As modern industry progressively evolves towards mass production, the complexity of material and information transfer in industrial processes has been significantly increasing [1]. This complexity is further amplified by the increasing automation of industrial systems and the widespread use of distributed control systems (DCS). As a result, data-based fault detection and diagnosis (FDD) methods have attracted considerable attention [2–4], as they can effectively identify the complex relationships hidden in large-scale industrial process data.

To effectively handle high-dimensional process data and achieve accurate process characterization, an increasing number of scholars have turned to statistical analysis-based methods to construct monitoring models. These methods include principal component analysis (PCA) [5], independent component analysis (ICA) [6], and support vector machines (SVM) [7].

However, these traditional methods increasingly demonstrate their limitations when confronted with complex processes, nonlinear characteristics, and other practical challenges. In contrast, deep learning methods exhibit significant advantages in handling complex relationships and can effectively learn intricate abstract features. Among deep learning-based fault identification methods, the autoencoder (AE) is the most representative. To take into account the dynamic and non-linear information between process samples, Li *et al.* [8] proposed a process monitoring scheme based on spatio-temporal neighborhood enhanced sparse autoencoder. Yang *et al.* [9] proposed a new stacked autoencoder (SAE) fault identification model to deal with adaptive fault identification problem. Convolutional neural networks (CNN) are also

* Corresponding author. East China University of Science and Technology, Shanghai 200237, China.
E-mail address: hbshi@ecust.edu.cn (H. Shi).

https://doi.org/10.1016/j.cjche.2025.10.002

the mainstream networks which has the advantage of capturing spatially localized dependencies compared to AE. Chen *et al.* [10] proposed one-dimensional convolutional autoencoder (1D-CAE) for noise reduction of high-dimensional process signals to learn hierarchical feature representations for fault identification. Zhang *et al.* [11] proposed an enhanced deep convolutional neural network (EDCNN) model for improved fault detection and diagnosis in chemical processes.

However, these methods indiscriminately stack variables in the input during feature extraction, neglecting the inherent topological information between process variables and the varying strengths of their interrelationships. This oversight makes it challenging to effectively aggregate information from variables that share similar mechanistic relationships during the feature extraction process. In fact, multiple variables captured by multiple sensors in spatial dimensions can naturally form graph structures in non-Euclidean spaces. Naturally, topology graphs are one of the commonly used structural models for such structured data [12–14], and studying the correlations between process variables from a graph-structured perspective has garnered widespread attention. Li *et al.* [15] established a GNN-based framework for intelligent fault identification and prediction and demonstrated significant advantages of the GNN in fault identification. Wu and Zhao [16] utilized the topological knowledge of the process to identify faults by connecting the process variables, manipulated variables, etc. in the control loop as nodes and constructing graph structures through priori mechanistic connections. Wu *et al.* [17] converted large-scale processes into directed graphs based on process knowledge and then decomposed them into multiple physically meaningful subblocks for fault identification. However, the correlations between the variable nodes in the actual process are different in strength magnitude. To take this problem into account, Chen and Ge [18] designed the spatial self-attention mechanism (SAM) to discover knowledge directly from the data and a graph-convolution layer to measure the relationships between process variables, both of which are combined to construct the soft measurement model. Deng *et al.* [19] combined the physical significance of the variables with the correlation analysis of the variables, and finally obtains the fault identification results through dynamically adjusting the graph structure. Most existing methods either rely solely on a rudimentary understanding of the process to construct topological models or employ purely data-driven approaches to learn node relationships. However, complex chemical processes involve numerous nonlinear subsystems that interact in intricate ways. Representing fault propagation information using only an incomplete physical topology or purely data mining methods is insufficient as these approaches overlook the guiding role of process topology.

Moreover, the inherent similarity in fault causes often results in comparable fault characteristics, which can easily lead to confusion and pose a significant challenge for accurate identification of similar faults. To address these issues, the sample-optimized dynamic convolutionally enhanced graph neural network (SOAP-EGNN) is proposed for large-scale industrial process fault identification. Initially, the construction of sensor relationships is guided by process mechanistic knowledge and compensatory causal topologies, enabling a multi-dimensional characterization of the relationships between process variables. Concurrently, the interdependencies between process variables are adaptively learned through edge-weight relationships in graph networks, thereby achieving adaptive perception of the varying degrees of influence that different faults exert on the relationships between variables. Ultimately, the sample-optimized feature processing strategy is employed to select representative samples for the identification task, ensuring that the extracted features are more discriminative.

The principal work and contributions of this research are encapsulated as follows.

(1) An innovative integration of industrial process knowledge with data is achieved, leveraging mechanistic knowledge and causal interactions to establish sensor relationships. By analyzing edge-weight relationships in topology graphs, dynamic interactions are addressed.
(2) An enhanced graph with dynamic convolution is designed to overcome static limitations, adaptively capturing temporal and topological relationships in process data for flexible and accurate analysis of complex systems.
(3) A sample-optimized feature processing strategy (SOFPS) is proposed to enhance fault identification, preserving distinct fault representations in manifold learning and maximizing inter-class separation, thereby improving identification performance for similar fault categories.

The remainder of this article is organized as follows: Section 2 provides a detailed description of the SOAP-EGNN fault identification method. Section 3 delves into the discussion of the essential experimental results. Finally, Section 4 presents the conclusions drawn from the study.

2. Sample-optimized Adaptive Perceptual Enhanced Graph Neural Network

For large-scale industrial process fault identification, a sample-optimized adaptive perceptual enhanced graph neural network is proposed. The overall framework of the network is depicted in Fig. 1, where the four modules in each part correspond to the four subsections that follow.

2.1. Strategic topology initialization through adjacency matrix analysis

In industrial processes, units are physically interconnected *via* pipelines, and process and manipulated variables are linked to sensors, controllers, actuators, and so on, through control loops. Consequently, process knowledge, such as PFD and P&ID, can be seamlessly transformed into a spatial topology graph. For process industry data, inputs to the graph model $\boldsymbol{X}_t \in \mathbb{R}^{n\times m}$ are derived from time series data using a sliding window of size m. The n variables serve as graph nodes. The time window sample at moment t is represented as follows:

$$\boldsymbol{X}_t = \begin{bmatrix} x_{1,t} & x_{1,(t-1)} & \cdots & x_{1,(t-m+1)} \\ x_{2,t} & x_{2,(t-1)} & \cdots & x_{2,(t-m+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,t} & x_{n,(t-1)} & \cdots & x_{n,(t-m+1)} \end{bmatrix} \tag{1}$$

From a messaging perspective, various sensor nodes articulate the process information conveyed by measurement elements for sensors and actuators, while directed edges delineate the direction of information transfer, representing the interdependencies among sensors and actuator measurement elements. To construct topology graphs with comprehensive information, rather than relying solely on incomplete mechanistic knowledge, the adjacency matrices of the topology graphs are supplemented by causal relationships between process variables, as depicted in Part 1 in Fig. 1. This approach ensures that information transfer and feature updating in graph networks depend solely on mechanically interrelated or causally connected variables. Suppose the adjacency matrix belonging to the *bth* subgraph $\boldsymbol{G}_b$ obtained from mechanistic knowledge and causal knowledge mining are $\boldsymbol{A}_{mb}$ and

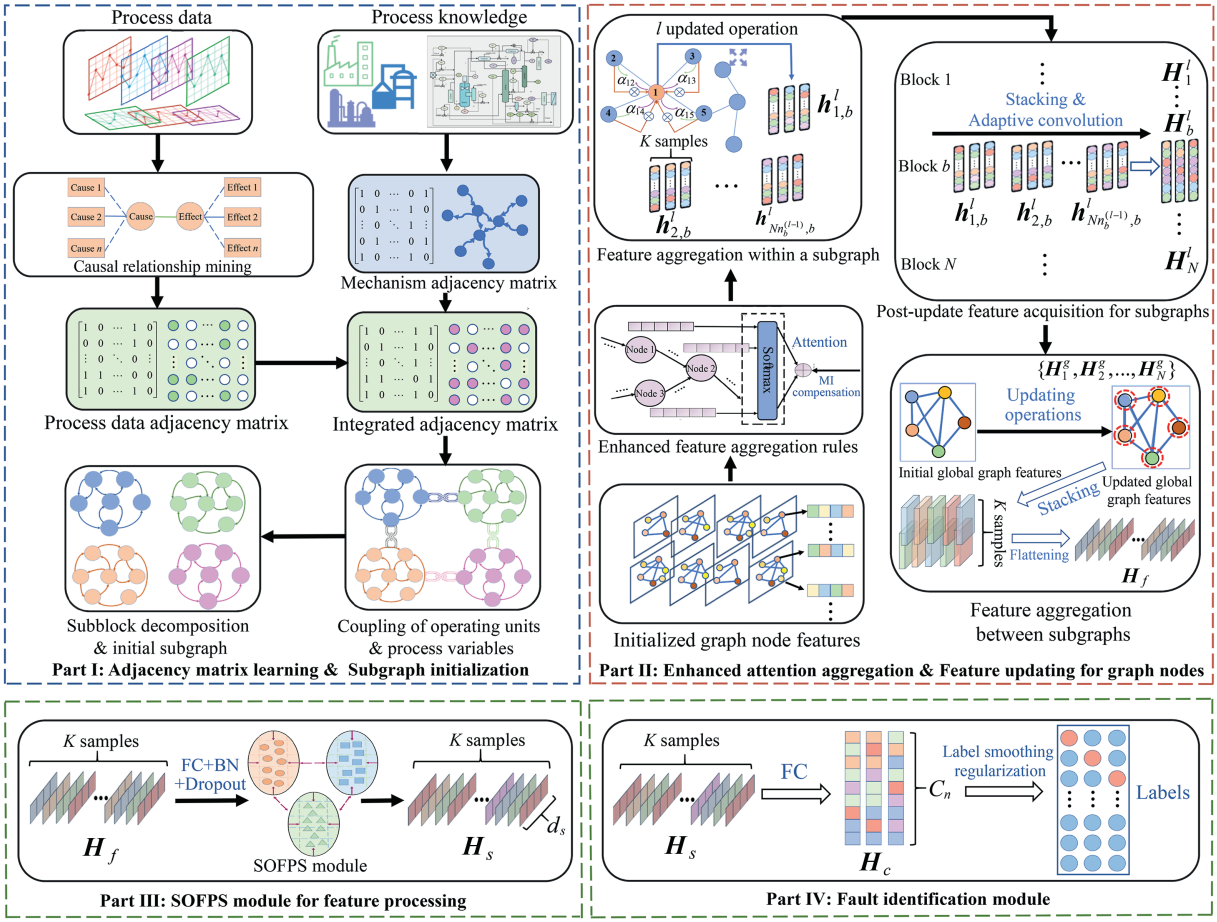


Fig. 1. Framework of the proposed SOAP-EGNN.

$\boldsymbol{A}_{cb}$, respectively. The implicit causal relationships between variables is often hidden in the process data and cannot be directly inferred from the production equipment and piping connections. The adjacency matrix $\boldsymbol{A}_{cb}$ is used as a medium to establish relationships between variable nodes, and the causality mechanism between variable nodes is introduced to initialize the causal adjacency matrix. The integrated adjacency matrix $\boldsymbol{A}_{lb}$ is obtained through the following decision criterion:

$$\boldsymbol{A}_{lb}^{ij} = \begin{cases} 1, \boldsymbol{A}_{mb}^{ij} = 1 \vee \boldsymbol{A}_{cb}^{ij} = 1 \\ 0, \boldsymbol{A}_{mb}^{ij} = 0 \wedge \boldsymbol{A}_{cb}^{ij} = 0 \end{cases} \tag{2}$$

where i, j represent the positional indices of row i and column j within the matrix. The collection of initial adjacency matrices for all subgraphs is characterized as follows:

$$\boldsymbol{A}_S = \boldsymbol{A}_{I1} \cup \boldsymbol{A}_{I2} \cup \ldots \cup \boldsymbol{A}_{IN} \tag{3}$$

where N denotes the number of subgraphs after subsystem decomposition. By considering variables as nodes, the industrial process can be abstractly represented as follows:

$$\textbf{Nodes} = \left\{ Node_{i,M\times D_j}, i \in (N_n, \ldots, 2, 1), j \in (D_n, \ldots, 2, 1) \right\} \tag{4}$$

where N_n denotes the total number of nodes in the node set, and D_j signifies the feature dimension of node i. Due to the mutual coupling among variables, anomaly information from a variable node is promptly relayed to neighboring nodes *via* the adjacency matrix, thus circumventing potential time delays associated with fault occurrences in the actual process. Based on the aforementioned set of nodes and the initialized adjacency matrices, the set of topology graphs $\boldsymbol{S}_{G_l}$ is initialized as follows:

$$\boldsymbol{S}_{G_l} = \boldsymbol{G}_{I1} \cup \boldsymbol{G}_{I2} \cup \ldots \cup \boldsymbol{G}_{Ib} \cup \ldots \cup \boldsymbol{G}_{IN} \tag{5}$$

2.2. Adaptive perceptual graph networks based feature aggregation and updating

2.2.1. Variable messaging based on enhanced attention mechanism and local adaptive perceptual layer at the subgraph variable level

Traditional adjacency matrices of topology graphs assign identical edge weights to all neighboring nodes, treating the entire process variables with uniform emphasis when integrating node feature information. However, this method fails to capture the strength or magnitude of sensor relationships. For instance, due to the intricate coupling among variables, a scenario may arise where variable A is linked to both B and C, yet a change in A significantly impacts B while minimally affecting C. This is a situation that qualitative sensor relationships cannot fully describe. Therefore, as illustrated in Part 2 of Fig. 1, the proposed method learns adaptive perceptual relationships from historical data. To ascertain the specific correlation strength between process variables, an

enhanced attention mechanism within graph neural networks is introduced:

$$\alpha_{ij}=\frac{\exp\left(\rho\left(\boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{W}\boldsymbol{h}_i^{(r)}\|\boldsymbol{W}\boldsymbol{h}_j^{(r)}\right]\right)\right)}{\sum\limits_{k\in N_i}\exp\left(\rho\left(\boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{W}\boldsymbol{h}_i^{(r)}\|\boldsymbol{W}\boldsymbol{h}_k^{(r)}\right]\right)\right)} \tag{6}$$

$$\boldsymbol{h}_i^{(r+1)}=\rho\left(\sum_{j\in N_i}\alpha_{ij}\boldsymbol{W}\boldsymbol{h}_j^{(r)}\right) \tag{7}$$

where an activation function $\rho(\cdot)$ is utilized to introduce non-linearity into the model. $\boldsymbol{h}_i^{(r)}$ denotes the features of node i in the rth layer, and $\{.\|.\}$ signifies the feature splicing operation.

To address the absence of prior relationships between nodes in traditional graph networks' attention mechanisms, enhanced aggregation and updating of subblock features are executed *via* a designed local adaptive perceptual layer at the variable level within each subgraph. Given the strong information transfer relationship among variables within the same operating unit, the initial unit topology is constructed using the aforementioned adjacency matrix learning based on process decomposition by operating equipment. Furthermore, the input $\boldsymbol{X}_t$ is segmented into $\left[\boldsymbol{X}^N,\ldots,\boldsymbol{X}^2,\boldsymbol{X}^1\right]^T$ based on process decomposition results, where $\boldsymbol{X}^b\in\mathbb{R}^{\mathrm{Num}_b\times m}$, and Num_b represents the number of variables in the bth block. Raw process measurements are encoded into hidden vectors as embedding features using the convolution operation:

$$\boldsymbol{H}_b^{(0)}=\boldsymbol{X}^b\boldsymbol{W}^b\in\mathbb{R}^{\mathrm{Num}_b\times d_0} \tag{8}$$

where $\left(\boldsymbol{H}_b^{(0)}\right)^{\mathrm{T}}=\left[\boldsymbol{h}_{1,b}^{(0)},\boldsymbol{h}_{2,b}^{(0)},\ldots,\boldsymbol{h}_{\mathrm{Num}_b,b}^{(0)}\right]$, $\boldsymbol{h}_{i,b}^{(0)}\in\mathbb{R}^{d_0\times 1}$ denotes the initial embedding feature of the ith node in the bth subblock, and d_0 denotes the dimensionality of the embedding features of each node. Subsequently, subblocks conduct information transmission and feature aggregation to effectively capture the interactions among variables:

$$\boldsymbol{g}_{i,b}^{(r+1)}=\rho\left(\widetilde{\boldsymbol{AT}}_{i,i}^{(r),b}\boldsymbol{h}_{i,b}^{(r)}+\sum_{j\in N_i}\widetilde{\boldsymbol{AT}}_{i,j}^{(r),b}\boldsymbol{h}_{j,b}^{(r)}\right)\in\mathbb{R}^{d_g^{(r+1)}} \tag{9}$$

$$\boldsymbol{h}_{i,b}^{(r+1)}=\rho\left(\widetilde{\boldsymbol{AT}}_{i,i}^{(r),b}\left(\boldsymbol{W}_{f_1}\boldsymbol{g}_{i,b}^{(r+1)}+b_{f1}\right)+\sum_{j\in N_i}\widetilde{\boldsymbol{AT}}_{i,j}^{(r),b}\left(\boldsymbol{W}_{f_2}\boldsymbol{g}_{j,b}^{(r)+1}+b_{f2}\right)\right)\in\mathbb{R}^{d_h^{(r+1)}} \tag{10}$$

where $\boldsymbol{h}_{i,b}^{(r)}$ denotes the node characteristics of the ith node in subblock b after the rth aggregated operation. The element $\widetilde{\boldsymbol{AT}}_{i,j}^{(r),b}$ represents the enhanced attention matrix element for the bth subblock after the rth operation. $\boldsymbol{W}_{f_1}$, $\boldsymbol{W}_{f_2}$ are learnable parameters, with b_{f_1} and b_{f_2} as bias terms. The representation of $\widetilde{\boldsymbol{AT}}^{(r),b}$ is divided into two levels: it manifests traditionally as a result of inter-feature attention coefficient calculations:

$$\left(\boldsymbol{AT}_{\mathrm{att}}\right)_{i,j}^{(r),b}=\begin{cases}\rho\left(\boldsymbol{a}^{(r),b}\left[\boldsymbol{h}_{i,b}^{(r)}\oplus\boldsymbol{h}_{j,b}^{(r)}\right]\right),\ \text{if}\ \boldsymbol{a}^{(r),b}\left[\boldsymbol{h}_{i,b}^{(r)}\oplus\boldsymbol{h}_{j,b}^{(r)}\right]>\mathrm{THR}^{(r)}\\ \delta,\ \text{if}\ \boldsymbol{a}^{(r),b}\left[\boldsymbol{h}_{i,b}^{(r)}\oplus\boldsymbol{h}_{j,b}^{(r)}\right]\le\mathrm{THR}^{(r)}\end{cases} \tag{11}$$

where $\mathrm{THR}^{(r)}$ denotes the screening threshold of the attention value information, representing the interaction of two nodes. The acquisition of node i's neighborhood N_i is facilitated by the aforementioned Eq. (2), and the direction of information transfer is from the neighbor feature node $\boldsymbol{h}_{j,b}^{(r)}$ to central feature node $\boldsymbol{h}_{i,b}^{(r)}$. To retain more relevant node relationships and perform graph information sparsification, redundant weakly relevant node relationships are discarded by assigning a comparatively minor value δ to them. Additionally, the mutual information between pairwise process variables is embedded into network learning, representing the strength of association between variables:

$$\left(\boldsymbol{AT}_{im}\right)^{(r),b}=\begin{cases}f_{\mathrm{top}k_r-\xi}\left(\begin{bmatrix}f_{\mathrm{qua}}(x_1,x_1) & f_{\mathrm{qua}}(x_1,x_2) & \cdots & f_{\mathrm{qua}}(x_1,x_{\mathrm{Num}_b})\\ f_{\mathrm{qua}}(x_2,x_1) & f_{\mathrm{qua}}(x_2,x_2) & \cdots & f_{\mathrm{qua}}(x_2,x_{\mathrm{Num}_b})\\ \vdots & \vdots & \ddots & \vdots\\ f_{\mathrm{qua}}(x_{\mathrm{Num}_b},x_1) & f_{\mathrm{qua}}(x_{\mathrm{Num}_b},x_2) & \cdots & f_{\mathrm{qua}}(x_{\mathrm{Num}_b},x_{\mathrm{Num}_b})\end{bmatrix}\right)\in\mathbb{R}^{\mathrm{Num}_b\times\mathrm{Num}_b},\ r=0\\ f_{\mathrm{top}k_r-\xi}\left(\begin{bmatrix}f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,1}^{(r)},\boldsymbol{c}_{b,1}^{(r)}\right) & f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,1}^{(r)},\boldsymbol{c}_{b,2}^{(r)}\right) & \cdots & f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,1}^{(r)},\boldsymbol{c}_{b,d_c^{(r)}}^{(r)}\right)\\ f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,2}^{(r)},\boldsymbol{c}_{b,1}^{(r)}\right) & f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,2}^{(r)},\boldsymbol{c}_{b,2}^{(r)}\right) & \cdots & f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,2}^{(r)},\boldsymbol{c}_{b,d_c^{r}}^{(r)}\right)\\ \vdots & \vdots & \ddots & \vdots\\ f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,d_c^{(r)}}^{(r)},\boldsymbol{c}_{b,1}^{(r)}\right) & f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,d_c^{(r)}}^{(r)},\boldsymbol{c}_{b,2}^{(r)}\right) & \cdots & f_{\mathrm{qua}}\left(\boldsymbol{c}_{b,d_c^{(r)}}^{(r)},\boldsymbol{c}_{b,d_c^{(r)}}^{(r)}\right)\end{bmatrix}\right)\in\mathbb{R}^{d_c^{(r)}\times d_c^{(r)}},\ r\in\mathbb{N}^*\end{cases} \tag{12}$$

where f_{qua} denotes a mutual-information-based quantification function, and the significance of $\boldsymbol{c}_{b,d_c^{(r)}}^{(r)}$ will be elaborated in Eq. (16). $f_{\text{top}k_r-\xi}(\cdot)$ retains the maximum k_r values per matrix row, with the retained variable count dynamically adjusted based on feature update count r. Discarded values are assigned a smaller value ξ. Mutual information reflects strong dependencies between feature pairs, supporting the method with prior relationships and forming a dual implicit representation alongside traditional attention calculations. Moreover, the original feature information remains uncorrupted by complex networks, incorporating invariant strongly-dependent prior information for each node. Thus, $\boldsymbol{AT}^{(r),b}$ can be expressed as follows:

$$\boldsymbol{AT}^{(r),b} = \boldsymbol{AT}_{\text{att}}{}^{(r),b} \odot \boldsymbol{AT}_{im}{}^{(r),b} \tag{13}$$

$$\widetilde{\boldsymbol{AT}}^{(r),b} = f_{\text{norm}}\left(\boldsymbol{AT}^{(r),b}\right) \tag{14}$$

where $\odot$ denotes the Hadamard product of matrices. Subsequently, each row of $\boldsymbol{AT}^{(r),b}$ is normalized using the function $f_{\text{norm}}(\cdot)$. In this study, the softmax function is employed to derive the final normalized enhanced attention matrix $\widetilde{\boldsymbol{AT}}^{(r),b}$. Subsequently, the updated features of each node within a block are aggregated to obtain the representative features of each block:

$$\boldsymbol{h}_b^{(r+1)} = f_{\text{stack}}\left(\boldsymbol{h}_{\text{Nn}_b^{(r)},b}^{(\text{r}+1)}, \ldots, \boldsymbol{h}_{2,b}^{(r+1)}, \boldsymbol{h}_{1,b}^{(r+1)}\right) \in \mathbb{R}^{Nn_b^{(r)} \times d_h^{(r+1)}} \tag{15}$$

where $Nn_b^{(r)}$ denotes the number of nodes at the end of the ith feature update operation and $d_h^{(r+1)}$ denotes the dimension corresponding to the current node features. Subsequently, subblock characteristics are refined according to Eq. (16) within sub-topology graphs, employing one-dimensional adaptive dynamic convolution (ADC) to extract block features from encoded latent vectors. The architecture of adaptive dynamic convolution is shown in Fig. 2, which dynamically adapts the shape and size of the convolutional kernel based on input data through a learnable deformation module:

$$\boldsymbol{c}_b^{(r+1)} = \sum_u \left(\theta_u * k_u^{(r+1)} \otimes \boldsymbol{W}_c \boldsymbol{h}_b^{(r+1)} + b^{(r+1)}\right) \in \mathbb{R}^{Cn^{(r+1)} \times d_c^{(r+1)}} \tag{16}$$

where $k_u^{(r+1)}$ represents the uth kernel of the ADC operation, $Cn^{(r+1)}$ signifies the number of output channels post one-dimensional ADC, and $d_c^{(r+1)}$ denotes the dimension of the output features after convolution. Besides, $\boldsymbol{h}_b^{(r+1)}$ is constructed by stacking feature nodes $\boldsymbol{h}_{1,b}^{(r+1)}$, $\boldsymbol{h}_{2,b}^{(r+1)}$, $\boldsymbol{h}_{3,b}^{(r+1)}$, and others, and they can be regarded as feature nodes within an abstract topological graph. θ_u denotes the weights assigned to different convolution kernels, which can be derived from the input data. $\boldsymbol{W}_c$ refers to learnable parameters, and $b^{(r+1)}$ indicates the bias term. For the dynamic convolution kernel parameters, they can be trained *via* backpropagation with gradient descent, aiming to minimize the loss function used during training, thereby learning and acquiring the specific parameters of the adaptive dynamic convolution kernels. This method allows for adaptive matching of the most suitable convolution kernels based on the input for feature extraction, thereby enhancing model performance.

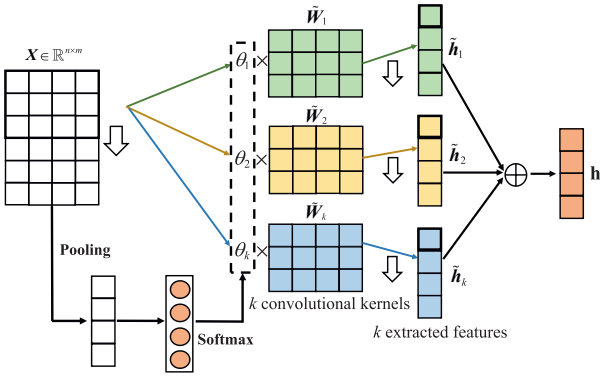


Fig. 2. Architecture of adaptive dynamic convolution.

Subsequently, a feature flattening operation is applied to $\boldsymbol{c}_b^{(r+1)}$, converting two-dimensional features into a one-dimensional vector for subsequent global feature aggregation and updates:

$$\boldsymbol{H}_b^{(r+1)} = f_{\text{flatten}}\left(\boldsymbol{c}_b^{(r+1)}\right) \in \mathbb{R}^{\left(Cn^{(r+1)} \times d_c^{(r+1)}\right)} \tag{17}$$

Given that the edge-weight relationships within the topology graph shift upon fault occurrence, with varying impacts across different segments of the industrial process, the scope and magnitude of these changes differ. Consequently, this leads to distinct fault feature categories when extracted using the topology graph built under normal conditions. The collaborative effect of the attention mechanism and ADC is particularly sensitive to weight variations between nodes. Ultimately, the sub-topology graphs, defined by adjacency matrices and quantitative edge weights, evolve in tandem with the progressive stages of feature aggregation:

$$\boldsymbol{S}_{G_{(r+1)}} = \boldsymbol{G}_{(r+1),N} \cup \ldots \cup \boldsymbol{G}_{(r+1),b} \cup \ldots \cup \boldsymbol{G}_{(\text{r}+1),2} \cup \boldsymbol{G}_{(\text{r}+1),1} \tag{18}$$

where $\boldsymbol{G}_{(r+1),b}$ denotes the bth subgraph following the $(r+1)$th graph update operation. The final aggregated features of the nodes associated with the N subblocks are derived through the l-step feature extraction process. A mapping relationship g is established between the updated subgraphs and their corresponding node features:

$$\boldsymbol{G}_{(l),b} = T_{(l)}^G\left(T_{(2)}^G\left(T_{(1)}^G\left(\boldsymbol{G}_{lb}\right)\right)\ldots\right) \tag{19}$$

$$\boldsymbol{H}_b^{(l)} = T_{(l)}^H\left(T_{(2)}^H\left(T_{(1)}^H\left(\boldsymbol{H}_b^{(0)}\right)\right)\ldots\right) \tag{20}$$

$$g : \left(\boldsymbol{G}_{(l),b} \in \boldsymbol{S}_{G_{(l)}}\right) \rightarrow \left(\boldsymbol{H}_b^{(l)} \in \left\{\boldsymbol{H}_N^{(l)}, \ldots, \boldsymbol{H}_2^{(l)}, \boldsymbol{H}_1^{(l)}\right\}\right) \tag{21}$$

where $T_{(i)}^G$ and $T_{(i)}^H$ denote the ith update operation performed on the topology graph structure and the subgraph aggregation features, respectively.

2.2.2. Global interactive perceptual layer at the sub-topology graph level

At the information interaction level of the sub-topology graph, global feature updates and acquisition are designed based on the proposed global interactive perceptual layer. Given the significant role of unit correlations in material transfer within industrial processes, and considering that simple concatenation of local aggregated features overlooks specific associations between operating units, we aim to uncover fine-grained coupling re-

lationships between cross-unit variables and facilitate global information interaction among all nodes. To this end, local feature representations are treated as new global-level feature nodes, forming the global topology $\boldsymbol{G}_{\text{global}}$, with feature updates designed accordingly:

$$\left(\boldsymbol{AT}_{\text{att,global}}\right)_{b,q} = \rho\left(\boldsymbol{a}_f\left[\boldsymbol{H}_b^{(l)} \middle\| \boldsymbol{H}_q^{(l)}\right]\right), q \in N_b \tag{22}$$

$$\widetilde{\boldsymbol{AT}}_{\text{att,global}} = f_{\text{norm}}\left(\boldsymbol{AT}_{\text{att,global}}\right) \tag{23}$$

$$\boldsymbol{F}_b = \rho\left(\left(\widetilde{\boldsymbol{AT}}_{\text{att,global}}\right)_{b,b} \boldsymbol{H}_b^{(l)} + \sum_{q \in N_b}\left(\widetilde{\boldsymbol{AT}}_{\text{att,global}}\right)_{b,q} \boldsymbol{H}_q^{(l)}\right) \in \mathbb{R}^{\left(Cn^{(l)} \times d_c^{(l)}\right)} \tag{24}$$

where $\boldsymbol{F}_b$ represents the output features associated with the bth node (subblock) following global aggregation and update operations. $\left(\boldsymbol{AT}_{\text{att,global}}\right)_{b,q}$ denotes the attention value between $\boldsymbol{H}_b^{(l)}$ and its neighboring node features $\boldsymbol{H}_q^{(l)}$, with f_{norm} defined as in Eq. (14). To manage excessive feature dimensions, a fully connected layer is employed for dimensionality reduction:

$$\boldsymbol{H}_b^g = \boldsymbol{W}_g \boldsymbol{F}_b + b_g \in \mathbb{R}^{d_h^g} \tag{25}$$

where d_h^g represents the dimension of the final aggregated features for each node, treating each subblock as an individual node at the global level. Ultimately, the global aggregated features, derived post feature update operation, encapsulate the entire chemical process and are utilized for the subsequent fault identification task:

$$\boldsymbol{H}_f = f_{\text{flatten}}\left(f_{\text{stack}}\left(\boldsymbol{H}_N^g, \ldots, \boldsymbol{H}_b^g, \ldots, \boldsymbol{H}_2^g, \boldsymbol{H}_1^g\right)\right) \in \mathbb{R}^{\left(N \times d_h^g\right)} \tag{26}$$

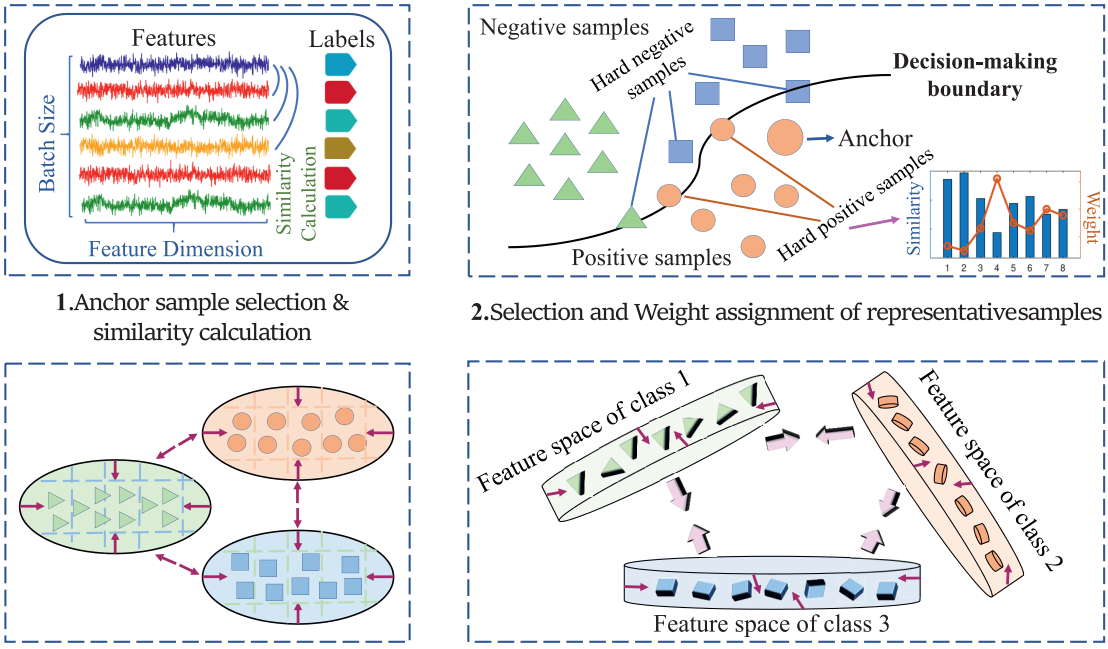
2.3. Sample-optimized feature processing strategy

Given the diversity of faults in actual industrial processes, some share similar causes and manifestations. Most fault identification methods struggle with distinguishing faults with analogous characteristics, primarily because they treat all samples uniformly without selective emphasis. Samples with valuable information for identifying similar faults are often eclipsed by redundant ones, diminishing the representativeness of the extracted features for fault identification. Thus, training a system capable of differentiating both distinctively and similarly characterized faults poses a challenge. The goal is to enhance intra-class compactness and inter-class separability within the low-dimensional manifold, ensuring each category has distinct decision boundaries:

$$\mathscr{L} = \max\left(0, d(f(s_a), f(s_p)) - d(f(s_a), f(s_n)) + \alpha\right) \tag{27}$$

where α is a marginal hyperparameter. $d(f(s_a), f(s_p))$ represents the distance between the anchor sample s_a and a positive s_p from the same category, and $d(f(s_a), f(s_n))$ represents the distance between the anchor sample s_a and a negative sample s_n from a different category. This function encourages the model to minimize the distance between samples of the same class (intra-class compactness) and maximize the distance between samples of different classes (inter-class separability).

To select representative samples near the classification boundaries that contain richer information and to optimize the attention given to different types of samples in the fault identification task, the sample-optimized feature processing strategy (SOFPS) is proposed, as illustrated in Part 3 of Fig. 1. The SOFPS



1.Anchor sample selection & similarity calculation

2.Selection and Weight assignment of representativesamples

3.Preserve within-class and between-class structure information in the low-dimensional feature space

4. Preserve within-class and between-class structure information in the high-dimensional feature space

Fig. 3. Procedures for sample-optimized feature processing strategy (SOFPS).

procedure, detailed in Fig. 3, initially segments batch feature samples into positive and negative pairs in the feature space using anchor samples and calculates inter-sample similarity. Subsequently, recognizing that many samples may hold limited valid information, a strategy for selecting representative samples and optimizing weight assignment is developed based on similarity calculations from the filtered samples. Hard (Representative) positive samples are those that closely resemble the anchor sample within the same category, while hard (representative) negative samples are those that are deceptively similar to the anchor sample but belong to different categories. By focusing on these samples, the model is forced to learn finer distinctions and more discriminative features because these representatives contribute more significantly to the gradients, leading to a more effective optimization process and improvement in the model's ability to distinguish between intra-class samples and inter-class samples.

If the similarity between a positive pair (same category as the anchor sample) and the anchor sample is nearly equivalent to that of a negative pair, it is labeled a hard positive pair. The concept of a hard negative pair is similarly defined. The sample-optimized feature processing steps are as follows. Assume the anchor sample is s_a, and a negative sample s_n^i is selected to meet the following equation:

$$\mathrm{Sim}_{a,i}^{-} + \varepsilon > \min_{y_k = y_a} \mathrm{Sim}_{a,k}^{+}, s_n^i \in \boldsymbol{S}_n, s_p^k \in \boldsymbol{S}_p \tag{28}$$

where $\mathrm{Sim}_{a,i}^{-}$ represents the similarity of the negative sample to the anchor sample, $\min_{y_k = y_a} \mathrm{Sim}_{a,k}^{+}$ denotes the minimum similarity of all positive samples to the anchor sample in a batch, and ε is a predefined threshold. If ε is too small, only a limited number of distinctively identifiable negative samples remain, potentially leading to incomplete sample coverage. Conversely, if ε is too large, redundant negative samples may be included, allowing unrepresentative samples to dominate the feature space. Hence, a moderate ε value is essential. $\boldsymbol{S}_n$ and $\boldsymbol{S}_p$ represent the sets of negative and positive samples related to the anchor sample, respectively. The selection criterion for a positive sample s_p^j is to meet the following condition:

$$\mathrm{Sim}_{a,j}^{+} - \varepsilon < \max_{y_k \neq y_a} \mathrm{Sim}_{a,k}^{-}, s_p^j \in \boldsymbol{S}_p, s_n^k \in \boldsymbol{S}_n \tag{29}$$

To handle the sparsity of samples, the mechanism that adaptively adjusts the influence of each sample based on its representativeness is introduced. This is achieved through the calculation of similarity scores and the application of a threshold to filter out redundant or less informative samples. To further address sample distribution and sparsity across categories, the method focuses on balanced representation within each fault category, maintaining proximity among same-class samples to prevent class imbalance and training bias. Additionally, it ensures each training batch includes representative samples from all classes in practical applications, keeping per-class sample counts relatively stable. This balanced strategy preserves model learning integrity and enhances fault identification robustness.

Following the aforementioned steps, the sets of representative negative samples $\boldsymbol{S}_{rn} = \{s_n^{a_m}, \ldots, s_n^{a_2}, s_n^{a1}\}$ and positive samples $\boldsymbol{S}_{rp} = \left\{s_p^{bm}, \ldots, s_p^{b_2}, s_p^{b_1}\right\}$ can be derived, where a_m and b_m denote the number of samples in the negative and positive sets, respectively. Subsequently, weight assignment for sample pairs is executed based on these candidate representative samples. For a chosen negative pair $\{s_n^i, s_p^j\} \in N$ or positive pair $\{s_p^i, s_p^j\} \in P$, the weight $w_{i,j}^{-}$ or $w_{i,j}^{+}$ is calculated as follows:

$$w_{ij}^{-} = \frac{1}{e^{\beta\left(\lambda - \mathrm{Sim}_{i,j}^{-}\right)} + \sum\limits_{k \in \boldsymbol{S}_{rn}} e^{\beta\left(\mathrm{Sim}_{i,k}^{-} - \mathrm{Sim}_{i,j}^{-}\right)}} = \frac{e^{\beta\left(\mathrm{Sim}_{i,j}^{-} - \lambda\right)}}{1 + \sum\limits_{k \in \boldsymbol{S}_{rn}} e^{\beta\left(\mathrm{Sim}_{i,k}^{-} - \lambda\right)}} \tag{30}$$

$$w_{ij}^{+} = \frac{1}{e^{-\alpha\left(\lambda - \mathrm{Sim}_{i,j}^{+}\right)} + \sum\limits_{k \in \boldsymbol{S}_{rp}} e^{-\alpha\left(\mathrm{Sim}_{i,k}^{+} - \mathrm{Sim}_{i,j}^{+}\right)}} = \frac{e^{-\alpha\left(\mathrm{Sim}_{i,j}^{+} - \lambda\right)}}{1 + \sum\limits_{k \in \boldsymbol{S}_{rp}} e^{-\alpha\left(\mathrm{Sim}_{i,k}^{+} - \lambda\right)}} \tag{31}$$

The SOFPS module enhances class separability by leveraging self-similarity and relative sample similarity. After processing through additional modules and the SOFPS, the feature $\boldsymbol{H}_s$ is derived from $\boldsymbol{H}_f$:

$$\boldsymbol{H}_s = f_{\mathrm{SOFPS}}\left[f_{\mathrm{BN}}\left(f_{\mathrm{Dropout}}\left(\boldsymbol{W}_s \boldsymbol{H}_f + b_s\right)\right)\right] \in \mathbb{R}^{d_s} \tag{32}$$

where f_{Dropout} represents the Dropout operation in neural networks, addressing overfitting in deep learning. f_{BN} signifies the batch normalization (BN) layer, where

$\boldsymbol{W}_s$, b_s are learnable parameters, and d_s denotes the feature dimensions of $\boldsymbol{H}_s$.

The SOFPS strategy effectively deals with feature overlap between different categories by enhancing intra-class compactness and inter-class separability within the low-dimensional manifold. This is achieved through a dynamic adjustment of weights based on the similarity of samples to the anchor sample. In summary, the separation of feature distribution in SOFPS is enforced through a dynamic and adaptive process that leverages the inherent structure of the data, ensuring that each category has distinct decision boundaries.

2.4. Label smoothing regularization based fault identification and overall loss function

Considering the nature of traditional classification tasks, the intermediate features extracted can be connected to subsequent layers to acquire task-specific features. As $\boldsymbol{H}_s$ remains a high-dimensional vector, direct application in downstream tasks with softmax is impractical. Thus, additional fault identification information is extracted from $\boldsymbol{H}_s$ using a fully connected feed-forward neural network to reduce dimensionality, as depicted in Part 4 in Fig. 1.

$$\boldsymbol{H}_c = \boldsymbol{W}_c \boldsymbol{H}_s + b_c \in \mathbb{R}^{C_n} \tag{33}$$

where C_n denotes the number of categories in the downstream task. The conventional cross-entropy loss initially computes softmax multiclass classification, mapping features from the sample $\boldsymbol{H}_c^i$ to each fault category *via* a fully connected layer. The predicted label for the current sample v_i is then determined using softmax regression as follows:

$$\widehat{l}_i = \underset{j}{\mathrm{argmax}} p_j(v_i) \frac{\exp\left(\sigma\left(\boldsymbol{W}_j^{\mathrm{T}} \boldsymbol{H}_c^i + b_j\right)\right)}{\sum\limits_{k=1}^{C_n} \exp\left(\sigma\left(\boldsymbol{W}_k^{\mathrm{T}} \boldsymbol{H}_c^i + b_k\right)\right)} \tag{34}$$

For the input samples $\boldsymbol{V}_{\mathrm{in}} = \{v_1, v_2, \ldots, v_{Sa_n}\}$, the cross-entropy loss is incorporated into the network's objective function:

$$L_{\text{cross}} = -\sum_{i=1}^{Sa_n}\sum_{j=1}^{C_n} q(j|v_i)\lg(p(j|v_i)) \tag{35}$$

where C_n denotes the total number of categories. $q(j|v_i)$ represents the true category label of the sample, and $p(j|v_i)$ indicates the output probability distributions across various patterns. Label smoothing regularization (LSR) is the process of making real labels less extreme by assigning them a certain fault-tolerant probability and assuming that labels have a probability of being mislabeled. This approach effectively relaxes the constraints on the label space by introducing a degree of uncertainty. To smooth the distribution of hard labels, the modified loss function incorporates soft labels generated by a uniform distribution:

$$q'(j|v_i) = (1-\varepsilon)q(j|v_i) + \frac{\varepsilon}{C_n} \tag{36}$$

$$L_{\text{slr}} = -\frac{1}{C_n}\sum_{j=1}^{C_n}\sum_{i=1}^{Sa_n} q'(j|v_i)\lg(p(j|v_i)) \tag{37}$$

where ε represents a minor constraint. Based on this, the loss function associated with input samples $\boldsymbol{V}_{\text{in}} = \{v_{Sa_n}, \ldots, v_2, v_1\}$, produced by SOFPS, is presented as follows:

$$L_{\text{fp}} = \frac{1}{Sa_n}\sum_{i=1}^{Sa_n}\left\{\frac{1}{\alpha}\lg\left[1+\sum_{k\in P_i} e^{-\alpha\left(\text{Sim}_{i,k}^{+}-\lambda\right)}\right] + \frac{1}{\beta}\lg\left[1 + \sum_{k\in N_i} e^{\beta\left(\text{Sim}_{i,k}^{-}-\lambda\right)}\right]\right\} \tag{38}$$

Consequently, the final loss function L_{overall} is a composite of L_{slr} and L_{fp}, which is detailed as follows:

$$L_{\text{overall}} = \gamma * L_{\text{slr}} + (1-\gamma) * L_{\text{fp}} \tag{39}$$

Minimizing the loss function *via* gradient descent and iteratively refining node features yields the optimal update count within subgraph l^*, resulting in the optimal subgraph configuration $\{\boldsymbol{G}_{\text{N}}^*, \ldots, \boldsymbol{G}_2^*, \boldsymbol{G}_1^*\}$ and ultimately determining the final global graph structure $\boldsymbol{G}_{\text{global}}^*$.

3. Simulation Experiments Study

3.1. Case design and parameters settings

In this section, the performance and effectiveness of the proposed SOAP-EGNN is demonstrated through Tennessee Eastman (TE) chemistry process, with revised TE simulations detailed in this paper [20]. The TE model is a realistic simulation for large chemical plants, as described in Refs. [21,22]. The certain faults, like fault 6, can halt the simulation due to process safety issues, resulting in a significantly smaller sample size and category imbalance. Faults 3, 9, and 15 are excluded due to minimal impact on mean, variance, and higher-order statistics, which are challenging to detect. Faults 16 and 21 are particularly elusive in the revised TE simulation, often misleading the identification of other faults. Thus, the fault identification experiment includes 15 fault data and normal process data. This paper selects 31 variables for fault identification, comprising 22 measurement variables and 9 manipulated variables, excluding constant sensors like valve position compressor re-cycle and agitators rotation, which are not involved in process control.

The model's learning efficacy and generalization capability are influenced by the learning rate and batch size, which dictate the weight iteration step size. Smaller batch sizes may result in insufficient fault category information, impairing fault identification performance, whereas larger batches can enhance accuracy but increase computational complexity. To ascertain the optimal learning rate and batch size, a grid search approach is employed, varying the learning rate from 0.001 to 0.1 and the batch size from 16 to 128. As depicted in Fig. 4 (a), the model achieves peak performance with a learning rate of 0.003 and a batch size of 64.

The proposed adaptive dynamic convolution is also integrated into the feature extraction of the SOAP-EGNN model. As shown in Fig. 4 (b), optimal fault identification is achieved with 4 convolution kernels of size 3. The values of several key parameters are summarized in Table 1.

Furthermore, the initial topology graph construction is depicted in Fig. 5, alongside a listing of variables associated with subgraph segmentation. To avoid overly complex mechanism relationships and increased spurious causality due to a large number of nodes within a subgraph, the number of nodes in a subgraph should be moderate. Specifically, the black edges in the

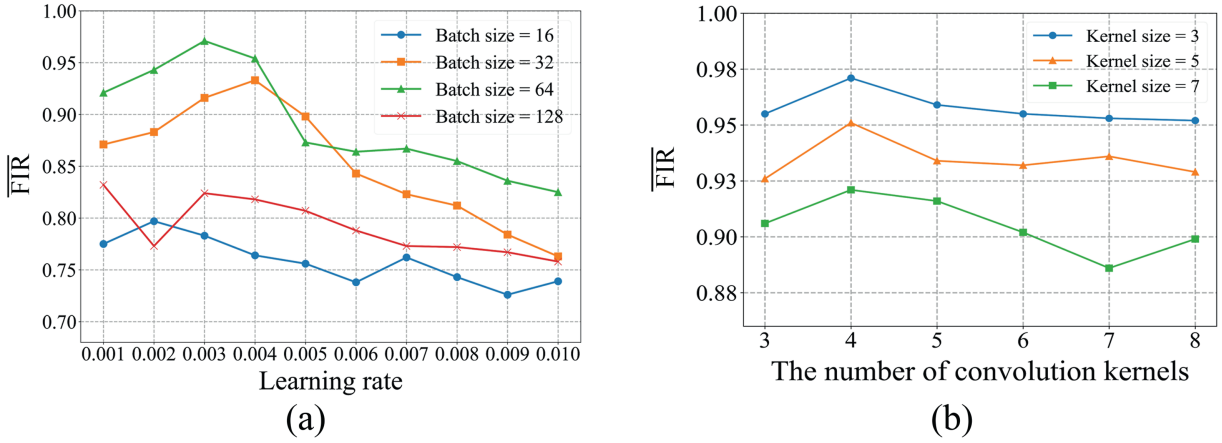


Fig. 4. Hyperparameter determination for network training. (a) Average FIR under different learning rates and batch sizes. (b) Average FIR under different number of convolution kernels and kernel sizes.

Table 1
Presentation of the values of parameters.

Names of the parameters	Values of the parameters
Learning rate	0.003
Batch size	64
The number of convolution kernels	4
The size of convolution kernels	3
The length of the time window T	20
The number of variables in the five units	8, 7, 6, 6, 4
The number of subblocks	5
The dimensionality of $\boldsymbol{H}_s$	$d_s = 128$
The total number of fault categories	$C_n = 16$
The number of feature aggregation and update operations of the subblocks	$l = 3$
The coefficient α	$\alpha = 2$
The coefficient β	$\beta = 40$
The coefficient λ	$\lambda = 0.5$
The smoothing coefficient ε	$\varepsilon = 0.1$
The balance coefficient in the overall loss function	$\gamma = 0.7$

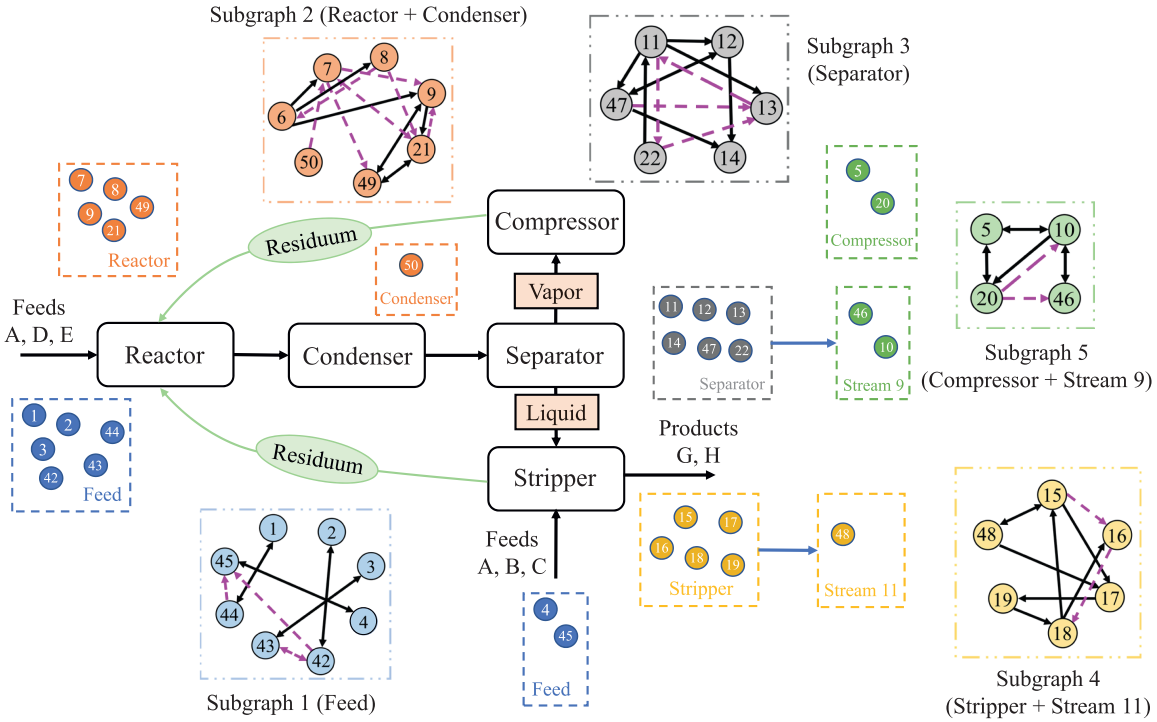


Fig. 5. Topology construction for operational units in the TE process.

graph are derived from process mechanistic knowledge, while edges sharing the same color as their respective nodes are determined through conditional Granger causality (CGC) analysis.

3.2. Comparative experiments

To substantiate the superior fault identification capabilities of SOAP-EGNN, it is compared with conventional deep learning techniques represented by MBTCN [23], D-SAE [24] as well as methods represented by the state-of-the-art graph-related structure GraphSAGE [25], GGCN [26], IAGNN [27], CTA-GNN [28] and OPNA-GNN [29]. The fault identification performance are quantified by fault identification rate (FIR), with confusion matrices and t-SNE visualizations for SOAP-EGNN and comparative methods presented in Fig. 6 and Fig. 7, respectively.

The t-SNE results indicate that SOAP-EGNN more effectively distinguishes between fault features compared to other methods, with less overlap and confusion. Table 2 compares FIR performance, highlighting the maximum FIR for each fault in bold. Notably, our method achieves an average FIR of 97.1%, as shown in the penultimate row of Table 2. The last row of Table 2 reveals that our approach not only attains a higher FIR across more fault categories but also excels in identifying the normal class, fault 8, fault 10, fault 11, fault 12, fault 13, and others, with 9 categories exceeding 99% accuracy. SOAP-EGNN's adaptive perceptual edge weights between nodes allow for a more precise detection of correlation changes caused by different faults, enhancing fault identification performance.

SOAP-EGNN surpasses existing methods by addressing their limitations in dynamic interaction handling, feature separation,

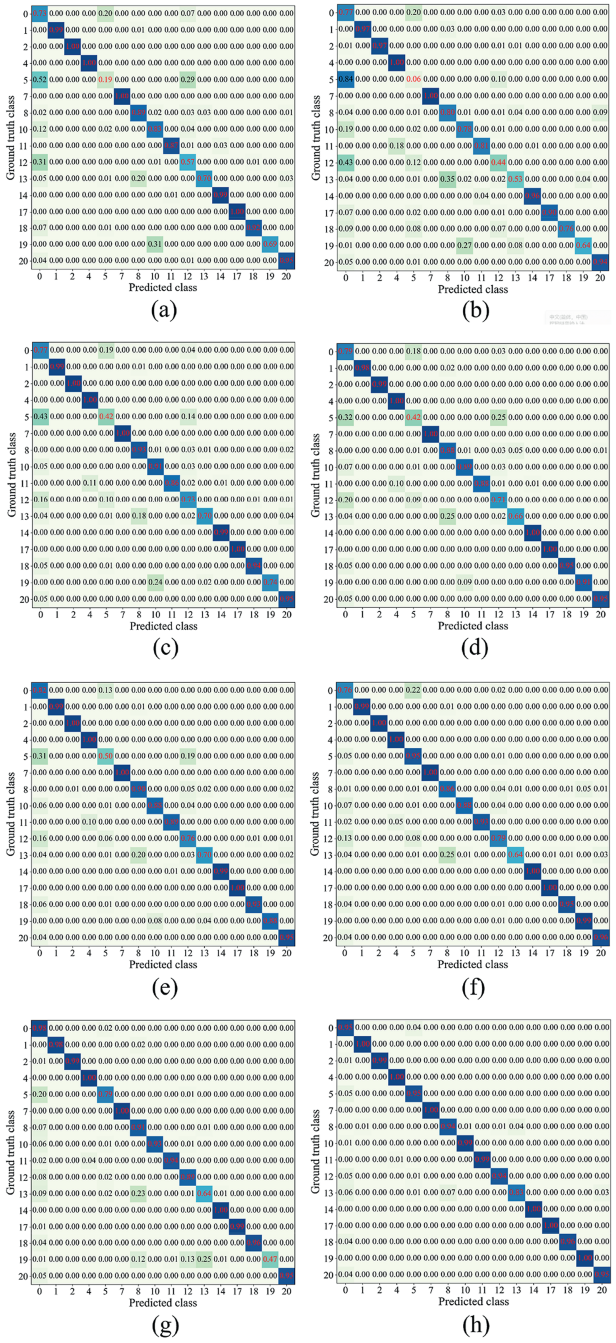


Fig. 6. Confusion matrices of SOAP-EGNN and other comparative methods in the TE process. (a) MBTCN. (b) D-SAE. (c) GraphSAGE. (d) GGCN. (e) IAGNN. (f) CTA-GNN. (g) OPNA-GNN. (h) SOAP-EGNN.

and causal relationship modeling. Unlike MBTCN, which focuses on temporal correlations, SOAP-EGNN integrates mechanistic knowledge and causal relationships for accurate sensor modeling. It outperforms D-SAE by using a sample-optimized strategy to enhance fault identification and maximize inter-class separation. While GraphSAGE relies on static graphs, SOAP-EGNN introduces dynamic convolution-embedded graphs to adaptively capture temporal and topological relationships. GGCN lacks mechanistic integration, whereas SOAP-EGNN bridges this gap through edge-weight analysis. IAGNN and CTA-GNN fall short in causal interactions and dynamic adaptation, which SOAP-EGNN resolves

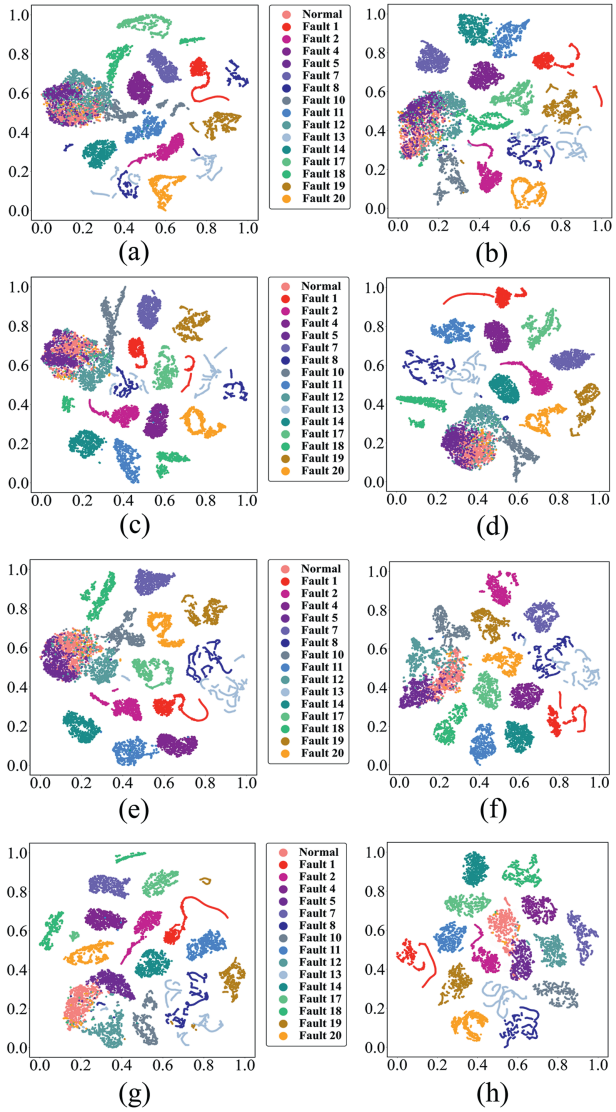


Fig. 7. t-SNE results of SOAP-EGNN and other comparative methods in the TE process. (a) MBTCN. (b) D-SAE. (c) GraphSAGE. (d) GGCN. (e) IAGNN. (f) CTA-GNN. (g) OPNA-GNN. (h) SOAP-EGNN.

with explicit causal modeling and enhanced graph structures. OPNA-GNN, while effective for gear faults, cannot match SOAP-EGNN's holistic integration of mechanistic knowledge, causal analysis, and dynamic modeling. Overall, SOAP-EGNN offers a superior solution for fault diagnosis in complex industrial systems.

3.3. Ablation experiments

In addition, five ablation studies are conducted, with the distinctions between the ablated models and SOAP-EGNN detailed in Table 3. The symbols $\surd$ and $\times$ denote the retention or omission of specific operations within the SOAP-EGNN framework, respectively. To substantiate the enhancement in fault identification performance afforded by our proposed strategies, comparative results are presented in Table 4. The data reveal a marked improvement in FIR upon implementing all strategies, particularly for the normal class, fault 5, fault 10, fault 11, fault 12, fault 13, and fault 19, as illustrated in Fig. 8.

Table 2
FIR of SOAP-EGNN and other comparative methods in the TE process.

	MBTCN	D-SAE	GraphSAGE	GGCN	IAGNN	CTA-GNN	OPNA-GNN	SOAP-EGNN
Normal	0.730	0.772	0.775	0.789	0.822	0.758	0.980	**0.955**
Fault 1	0.986	0.968	0.992	0.985	0.986	0.992	0.985	**0.997**
Fault 2	**0.997**	0.973	**0.997**	0.994	**0.997**	0.997	0.986	0.995
Fault 4	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Fault 5	0.188	0.058	0.423	0.425	0.501	0.945	0.792	0.946
Fault 7	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Fault 8	0.886	0.798	0.930	0.882	0.900	0.857	0.913	**0.939**
Fault 10	0.827	0.777	0.908	0.893	0.881	0.880	0.926	**0.992**
Fault 11	0.868	0.811	0.862	0.882	0.886	0.926	0.939	**0.992**
Fault 12	0.573	0.444	0.726	0.707	0.762	0.788	0.894	**0.942**
Fault 13	0.703	0.526	0.705	0.664	0.701	0.641	0.638	**0.831**
Fault 14	0.995	0.960	0.995	0.999	0.992	**1.000**	**1.000**	**1.000**
Fault 17	0.996	0.900	**1.000**	0.999	0.996	**1.000**	0.995	0.996
Fault 18	0.922	0.761	0.944	0.945	0.927	0.951	0.958	**0.960**
Fault 19	0.685	0.637	0.742	0.912	0.882	0.995	0.471	**1.000**
Fault 20	0.953	0.941	0.953	0.953	0.954	**0.960**	0.946	0.950
Average	0.831	0.770	0.871	0.876	0.886	0.916	0.902	**0.971**
Best times	3	2	4	2	3	5	3	**12**

Note: (1) Average: Average fault identification rate for all faults.
(2) Best times: Number of times each method has the highest individual fault identification rate among all methods.

Table 3
Experimental setup for ablation experiments under comparison with SOAP-EGNN.

	Block segmentation	Dynamic Convolution	Subgraph aggregation	MI enhanced	Causality compensation	LSR
Model 1	×	✓	✓	✓	✓	✓
Model 2	✓	×	✓	✓	✓	✓
Model 3	✓	✓	×	✓	✓	✓
Model 4	✓	✓	✓	×	✓	✓
Model 5	✓	✓	✓	✓	×	✓
Model 6	✓	✓	✓	✓	✓	×

Table 4
FIR of SOAP-EGNN and other ablation methods in the TE process.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SOAP-EGNN
Normal	0.674	0.801	0.814	0.937	0.823	0.825	**0.955**
Fault 1	0.991	0.988	0.985	0.982	0.985	0.965	**0.997**
Fault 2	**1.000**	0.996	0.992	0.996	0.995	0.987	0.995
Fault 4	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Fault 5	0.412	0.707	0.524	0.566	0.927	0.451	**0.946**
Fault 7	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Fault 8	0.904	0.893	**0.944**	0.910	0.886	0.811	0.939
Fault 10	0.896	0.894	0.895	0.957	0.928	0.931	**0.992**
Fault 11	0.840	0.887	0.907	0.880	0.919	0.907	**0.992**
Fault 12	0.662	0.817	0.632	0.642	0.730	0.932	**0.942**
Fault 13	0.688	0.702	0.748	0.697	0.785	0.572	**0.831**
Fault 14	0.990	0.997	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Fault 17	0.999	**1.000**	**1.000**	**1.000**	0.996	0.981	0.996
Fault 18	0.940	0.930	0.936	0.958	0.950	0.935	**0.960**
Fault 19	0.835	0.918	0.775	0.827	0.636	0.987	**1.000**
Fault 20	0.951	0.951	0.950	**0.955**	0.954	0.941	0.950
Average	0.859	0.904	0.881	0.895	0.906	0.886	**0.971**
Best times	3	3	5	5	3	3	**12**

3.4. Further comparison with state-of-art feature processing strategies and regularization methods

The commonly employed feature processing methods (FPMs) primarily focus on intra-class feature aggregation and inter-class feature separation through loss functions like N-pair-ms loss [30], Supcon loss [31], Circle loss [32], IAAM loss [33] and Perceptual loss [34]. To demonstrate the impact of various FPMs on fault identification performance, comparative experiments are conducted while maintaining the model's fundamental structure. The average FIR across different FPMs is depicted in Fig. 9, highlighting that our proposed FPM achieves an FIR of 97.1%.

Based on prior research experience, certain fault classes, specifically faults 5, 8, 10, 12 and 13, are intrinsically difficult to discriminate. These faults exhibit pronounced multidimensional similarity, which markedly increases their mutual confusability during classification and diagnosis. Faults 5 and 12 both originate from anomalous inlet temperatures in the condenser and display nearly identical thermal transients. Faults 8, 10 and 12 are characterized by stochastic perturbations whose signal dynamics

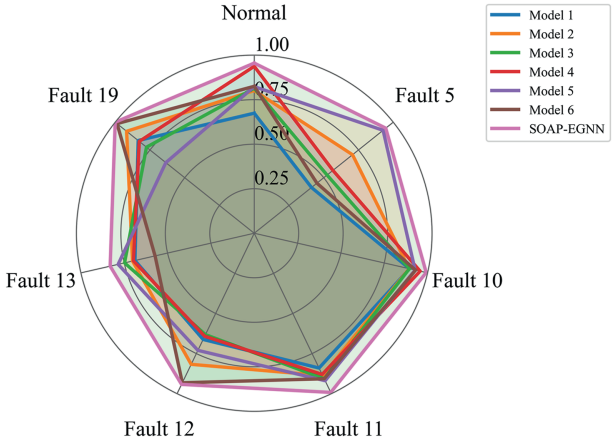


Fig. 8. FIR comparisons of ablation methods in the TE process.

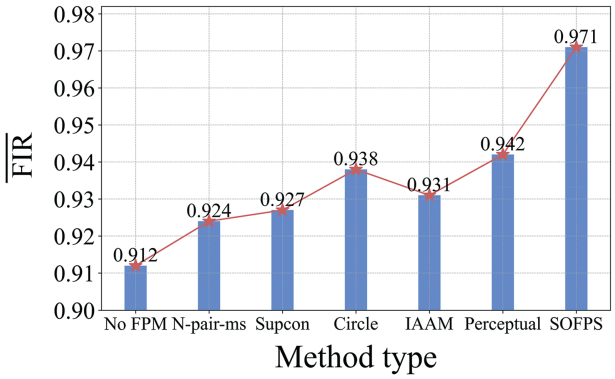


Fig. 9. Average FIR comparisons under different FPMs in the TE process.

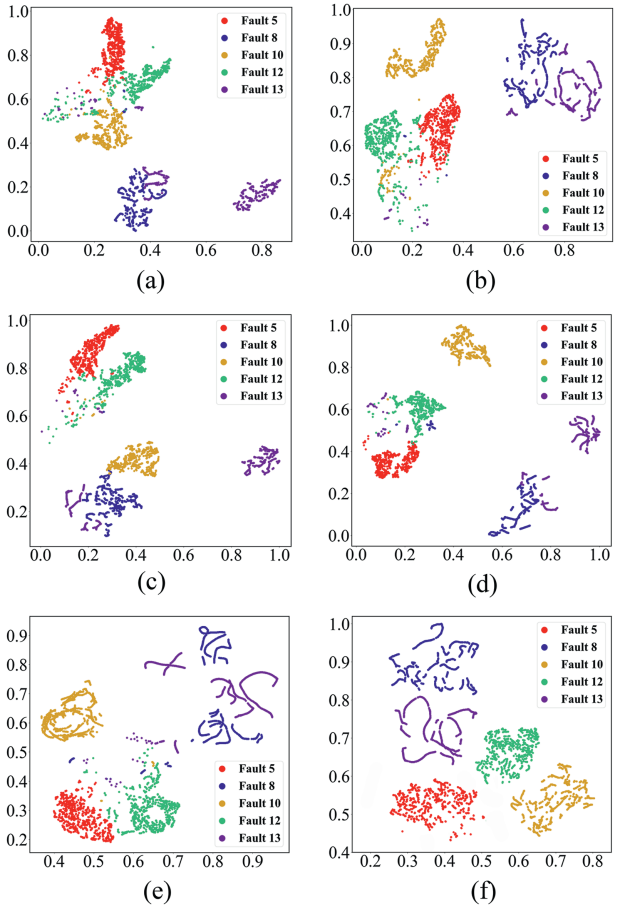


Fig. 10. t-SNE results of faults 5, 8, 10, 12, 13 in the TE process. (a) With N-pair-ms loss. (b) With Supcon loss. (c) With Circle loss. (d) With IAAM loss. (e) With Perpectual loss. (f) With SOFPS.

closely resemble normal operational variability. Additionally, faults 10, 12 and 13 manifest as slowly evolving state drifts, further aggravating pairwise similarity. Collectively, these overlapping manifestations challenge the separability of the aforementioned fault types. Moreover, these faults are all subject to multivariable coupling effects, which cause the system to exhibit highly similar abnormal patterns. Consequently, they share commonalities in variable fluctuation, evolution mechanisms, and coupling effects, forming a fault set with high mutual confusion. This effectively reflects the classification challenges posed by complex anomalies in real industrial processes.

Fig. 10 (a) illustrates that under the N-pair-ms loss, the feature distributions of the five fault types still exhibit significant overlap. In particular, faults 5, 10, and 12 share overlapping regions in the feature space, while faults 8 and 13 also demonstrate partial mixing, indicating that the inter-class boundaries remain indistinct. As shown in Fig. 10 (b), incorporating the Supcon loss results in clearer separation between faults 8 and 13; however, the features of faults 5, 10, and 12 continue to overlap in the feature space, suggesting that fault discrimination remains challenging. Fig. 10 (c) presents the results obtained with the Circle loss, where regulating intra-class and inter-class distances enhances discriminability to some extent, leading to relatively compact clusters for most categories. Nonetheless, partial confusion among certain fault features remains observable. In Fig. 10 (d), the IAAM loss fails to effectively enlarge inter-class separations, with several clusters still exhibiting overlap, indicating its limitations in handling complex multi-class fault recognition. Furthermore, Fig. 10 (e) illustrates the feature distribution under the Perceptual loss.

Table 5
FIR of SOAP-EGNN with LSR strategy and other regularization methods in the TE process.

	L2 regularization	OGLL-L1	MC-dropout	LSR strategy
Normal	0.893	0.881	0.923	0.955
Fault 1	0.962	0.977	0.995	0.997
Fault 2	0.997	0.991	1.000	0.995
Fault 4	1.000	1.000	1.000	1.000
Fault 5	0.926	0.678	0.867	0.946
Fault 7	1.000	0.999	1.000	1.000
Fault 8	0.834	0.866	0.923	0.939
Fault 10	0.971	0.939	0.932	0.992
Fault 11	0.889	0.953	0.997	0.992
Fault 12	0.919	0.912	0.781	0.942
Fault 13	0.734	0.683	0.671	0.831
Fault 14	1.000	1.000	1.000	1.000
Fault 17	0.999	0.986	1.000	0.996
Fault 18	0.949	0.928	0.959	0.960
Fault 19	0.774	0.991	1.000	1.000
Fault 20	0.955	0.940	0.959	0.950
Average	0.912	0.919	0.938	0.971

Although this method improves feature extraction to a certain degree, the boundaries between faults 5 and 12, as well as between faults 8 and 13, remain insufficiently distinct, suggesting constrained feature disentanglement capability. In contrast, as shown in Fig. 10 (f), SOFPS achieves clear boundaries and low-overlap feature distributions, enabling effective inter-class separation while preserving intra-class consistency, thereby providing

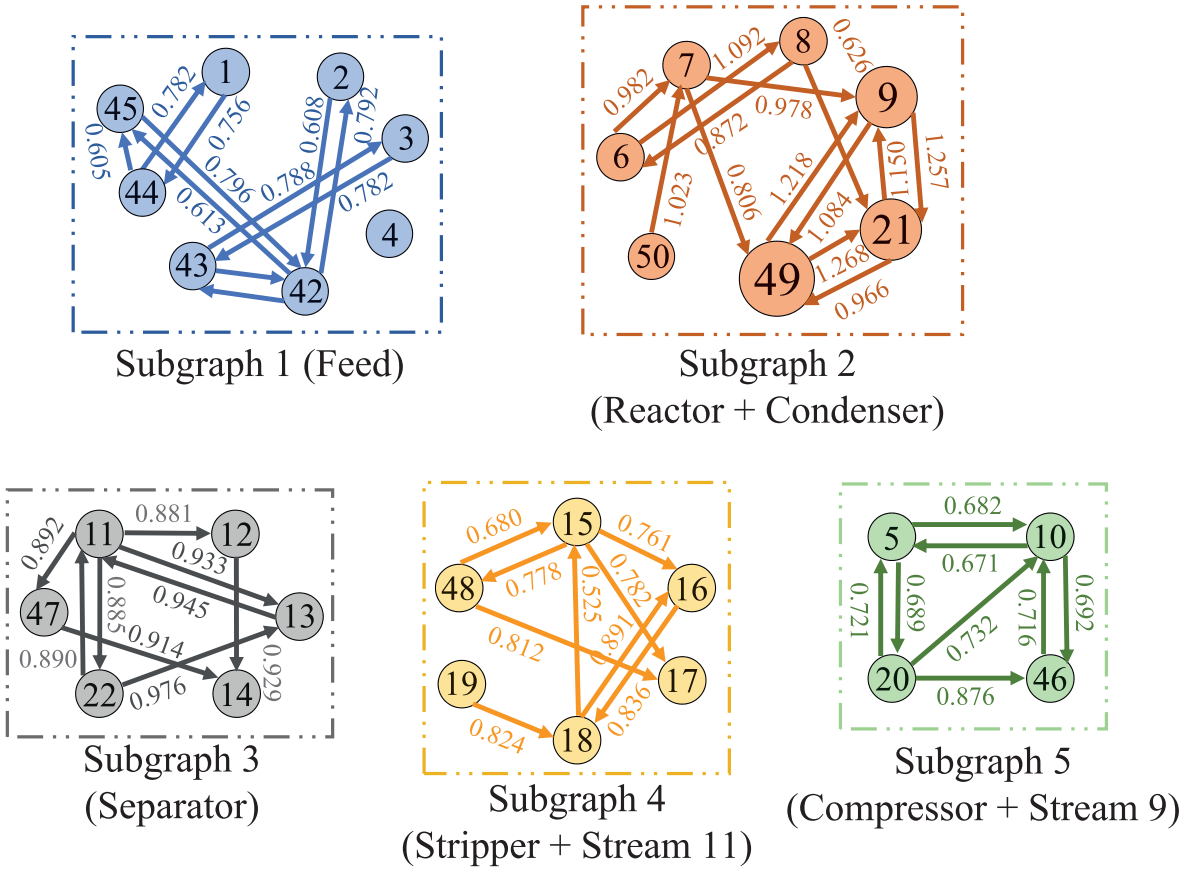


Fig. 11. Initial enhanced attention values of fault 4 state in the TE process.

more discriminative feature representations for subsequent fault identification.

Furthermore, the results presented in Table 5 provide a comprehensive evaluation of the SOAP-EGNN model under different regularization strategies in the TE process. The analysis highlights the superiority of the label smoothing regularization (LSR) strategy over comparative regularization methods such as L2 regularization [35], OGLL-L1 regularization [36], and MC-dropout [37]. The FIR values in Table 5 demonstrate that the LSR strategy consistently outperforms other regularization methods across most fault scenarios. SOAP-EGNN with LSR achieves an average FIR of 0.971, surpassing L2 regularization (0.912), OGLL-L1 regularization (0.919), and MC-dropout (0.938). This superiority underscores the effectiveness of LSR in enhancing model robustness and generalization. Notably, LSR achieves perfect performance in several fault cases such as fault 4, fault 7, fault 14, and fault 19, while other methods struggle in some of these scenarios. For example, L2 regularization and OGLL-L1 regularization fail to achieve perfect identification accuracy in fault 19, whereas LSR excels. The experimental results conclusively validate the superiority of the LSR strategy in enhancing SOAP-EGNN's fault identification capabilities.

3.5. Specific implementation results of enhanced attention mechanism

The proposed dual-attention design integrates prior relationships with data-driven feature learning, enabling the model to effectively capture complex node interactions while maintaining robustness. To further elaborate the implementation details of the enhanced attention mechanism and illustrate the differences in attention weights among graph nodes under the fault condition, an analysis of TE process fault 4 has been provided. As illustrated in Fig. 11, larger node sizes indicate a higher central importance of the node, reflecting increased attention values following a fault occurrence. Fault 4 is a step fault, which is caused by a change in the reactor coolant inlet temperature. As presented in Fig. 11, the attention values associated with V_{49} (recycle coolant flow rate), V_{21} (recycle coolant temperature), and V_9 (reactor temperature) become significantly more pronounced. This pinpoints the significance of variables like V_{49}, V_{21}, and V_9, which are directly affected by changes in the reactor coolant inlet temperature.

The enhanced attention mechanism thus provides a powerful tool for diagnosing and understanding complex fault scenarios in the TE process by adaptively focusing on relevant variables and their interactions, which also enhances interpretability by providing a clear visualization of the model's decision-making process, making it easier to understand the relationships and interactions between variables.

4. Conclusions

In this article, a novel framework termed SOAP-EGNN is designed for fault identification in large-scale processes. To investigate the interdependencies among process variables and how fault occurrences impact these relationships, graph neural networks are adopted as the foundational structure. Process data mining is leveraged to uncover sensor relationships not fully captured by incomplete process mechanistic knowledge, enhancing the description of

sensor connections. Subsequently, attention-based dynamic convolution models fine-grained quantitative dependencies between process variables, considering the prior relevance of variables, which facilitates adaptive fault perception. The integration of a sample-optimized feature processing strategy within SOAP-EGNN shifts decision boundaries towards low-density regions, clarifying distinctions among fault categories and enhancing fault identification performance. Additionally, the label smoothing regularization (LSR) strategy is employed to prevent overfitting during training. Experiments on the TE process demonstrate SOAP-EGNN's superiority over alternative methods for process fault identification, with ablation studies confirming the significant performance improvement of the fault identification.

Despite the innovative approach of SOAP-EGNN in enhancing fault identification through advanced graph neural networks and feature processing strategies, the current study has limitations in generalizability and scalability. Moreover, due to computational efficiency concerns, the proposed method may not be able to fully address the complexity of real-time fault detection. Future work should focus on expanding the model's validation to diverse industrial scenarios, optimizing computational efficiency for real-time applications, and exploring accurate identification of unknown faults.

CRediT Authorship Contribution Statement

Keyu Yao: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Hongbo Shi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation. Bing Song: Software, Funding acquisition. Yang Tao: Software, Funding acquisition.

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

This research is sponsored by the National Natural Science Foundation of China (62473154, 62473155, 62473156), the Natural Science Foundation of Shanghai (19ZR1473200), Shanghai Oriental Talents Program Youth Project (QNKJ2024034), Shanghai Chenguang Project (21CGA37).

References

[1] W.D. Tian, J.W. Zhang, Z. Cui, H.R. Zhang, B. Liu, Microscopic mechanism study and process optimization of dimethyl carbonate production coupled biomass chemical looping gasification system, *Chin. J. Chem. Eng.* 58 (2023) 291–305.
[2] F.N. Irani, M. Yadegar, N. Meskin, Koopman-based deep iISS bilinear parity approach for data-driven fault diagnosis: experimental demonstration using three-tank system, *Control Eng. Pract.* 142 (2024) 105744.
[3] 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.
[4] C.H. Zhao, Perspectives on nonstationary process monitoring in the era of industrial artificial intelligence, *J. Process Control* 116 (2022) 255–272.
[5] W.D. Tian, G.X. Zhang, X. Zhang, Y.X. Dong, PCA weight and Johnson transformation based alarm threshold optimization in chemical processes, *Chin. J. Chem. Eng.* 26 (2018) 1653–1661.
[6] A.N. M, S. R, Combined classification models for bearing fault diagnosis with improved ICA and MFCC feature set, *Adv. Eng. Software* 173 (2022) 103249.
[7] K.X. Bi, T. Qiu, An intelligent SVM modeling process for crude oil properties prediction based on a hybrid GA-PSO method, *Chin. J. Chem. Eng.* 27 (2019) 1888–1894.
[8] N.X. Li, H.B. Shi, B. Song, Y. Tao, Temporal-spatial neighborhood enhanced sparse autoencoder for nonlinear dynamic process monitoring, *Processes* 8 (2020) 1079.
[9] C. Yang, Z. Lai, Y.C. Wang, S.L. Lan, L.H. Wang, L.H. Zhu, A novel bearing fault diagnosis method based on stacked autoencoder and end-edge collaboration, in: 2023 26th Int. Conf. Comput. Support. Coop. Work Des. CSCWD, IEEE, Rio de Janeiro, Brazil, 2023.
[10] S.M. Chen, J.B. Yu, S.J. Wang, One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes, *J. Process Control* 87 (2020) 54–67.
[11] J.X. Zhang, M. Zhang, Z.M. Feng, R.F. Lv, C.Y. Lu, Y.Y. Dai, L.C. Dong, Gated recurrent unit-enhanced deep convolutional neural network for real-time industrial process fault diagnosis, *Process Saf. Environ. Prot.* 175 (2023) 129–149.
[12] Z.H. Wu, S.R. Pan, F.W. Chen, G.D. Long, C.Q. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, *IEEE Transact. Neural Networks Learn. Syst.* 32 (2021) 4–24.
[13] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, H. Liu, Graph learning: a survey, *IEEE Trans. Artif. Intell.* 2 (2021) 109–127.
[14] C.Y. Yang, J. Liu, K.B. Zhou, X.H. Yuan, M.-F. Ge, Transfer graph-driven rotating machinery diagnosis considering cross-domain relationship construction, *IEEE ASME Trans. Mechatron.* 27 (2022) 5351–5360.
[15] T.F. Li, Z. Zhou, S.N. Li, C. Sun, R.Q. Yan, X.F. Chen, The emerging graph neural networks for intelligent fault diagnostics and prognostics: a guideline and a benchmark study, *Mech. Syst. Signal Process.* 168 (2022) 108653.
[16] D. Wu, J.S. Zhao, Process topology convolutional network model for chemical process fault diagnosis, process, *Saf. Environ. Prot.* 150 (2021) 93–109.
[17] W.Q. Wu, C.Y. Song, J. Liu, J. Zhao, Data-knowledge-driven distributed monitoring for large-scale processes based on digraph, *J. Process Control* 109 (2022) 60–73.
[18] Z.C. Chen, Z.Q. Ge, Knowledge automation through graph mining, convolution, and explanation framework: a soft sensor practice, *IEEE Trans. Ind. Inf.* 18 (9) (2021) 6068–6078.
[19] Q. Deng, Z.W. Chen, P. Tang, X.H. Li, C.H. Yang, X. Yang, A novel semi-supervised fault diagnosis method for chillers based on neighbor-optimized graph convolutional network, *Energy Build.* 301 (2023) 113703.
[20] A. Bathelt, N.L. Ricker, M. Jelali, Revision of the Tennessee eastman process model, *IFAC-PapersOnLine* 48 (2015) 309–314.
[21] S. Yin, S.X. Ding, A. Haghani, H.Y. Hao, P. Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process, *J. Process Control* 22 (9) (2012) 1567–1581.
[22] Q.C. Jiang, X.F. Yan, B. Huang, Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and bayesian inference, *IEEE Trans. Ind. Electron.* 63 (2016).
[23] Y.M. He, H.B. Shi, S. Tan, J.Z. Zhu, Multiblock temporal convolution network-based temporal-correlated feature learning for fault diagnosis of multivariate processes, *J. Taiwan Inst. Chem. Eng.* 122 (2021) 78–84.
[24] S.Y. Liu, J. He, Z.W. Chen, D.F. Chen, Y. Chen, Discriminative stacked autoencoder: feature-integration boosting for bearing fault diagnosis, *IEEE Sens. J.* 23 (2023) 27549–27558.
[25] S.P. Gao, Y.Y. Li, D. Zhao, Fault diagnosis for rolling bearings based on novel visibility graph and GCN scheme, in: 2023 IEEE 12th Data Driven Control Learn. Syst. Conf. DDCLS, IEEE, Xiangtan, China, 2023.
[26] Y.D. Xu, J.C. Ji, Q. Ni, K. Feng, M. Beer, H.T. Chen, A graph-guided collaborative convolutional neural network for fault diagnosis of electromechanical systems, *Mech. Syst. Signal Process.* 200 (2023) 110609.
[27] D.Y. Chen, R.N. Liu, Q.H. Hu, S.X. Ding, Interaction-aware graph neural networks for fault diagnosis of complex industrial processes, *IEEE Transact. Neural Networks Learn. Syst.* 34 (9) (2021) 6015–6028.
[28] H. Wang, R.N. Liu, Z.X. Li, H.K. Zhou, Causal-trivial attention graph neural network for fault diagnosis of complex industrial processes, *IEEE Trans. Ind. Inf.* 20 (2) (2023) 1987–1996.
[29] B. Wang, Y.D. Xu, M.Y. Wang, Y.Z. Li, Gear fault diagnosis method based on the optimized graph neural networks, *IEEE Trans. Instrum. Meas.* 73 (2023) 1–11.
[30] T.-N. Hoang, D. Kim, Supervised contrastive ResNet and transfer learning for the in-vehicle intrusion detection system, *Expert Syst. Appl.* 238 (2024) 122181.
[31] J. Wan, D.X. Zhong, H.K. Shao, Palmprint recognition system for mobile device based on circle loss, *Displays* 73 (2022) 102214.
[32] C.C. Fan, Y.X. Wang, Y.H. Zhang, W.L. Ouyang, Interpretable multi-scale neural network for granger causality discovery, in: ICASSP 2023-2023 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP, IEEE, Rhodes Island, Greece, 2023.
[33] R.C. Zhi, Y.C. Meng, J.Y. Hou, J. Wan, Dual balanced class-incremental learning with im-Softmax and angular rectification, *IEEE Transact. Neural Networks Learn. Syst.* 36 (3) (2025) 4437–4447.

[34] X.C. Zhan, A.M. Li, J. He, L. Xiang, Enhancing low-dose pet imaging: a novel contrastive learning method for perceptual loss and an organ-aware loss, in: 2024 IEEE International Symposium on Biomedical Imaging, IEEE, 2024.
[35] K.A. Cahyanto, M. Yani, A. Kustiyo, The effects of L2 regularization in EfficientNet for human skin disease multi-class classification, in: 2023 6th International Conference of Computer and Informatics Engineering, IEEE, 2023.
[36] Z. Liu, H.J. Gu, Z. Chen, B. Pierre, Y.J. Yang, Dual-modal image reconstruction for electrical impedance tomography with overlapping group lasso and Laplacian regularization, *IEEE Trans. Biomed. Eng.* 70 (8) (2023) 2362–2373.
[37] X.X. Liu, C.C. Zou, C. Peng, C.W. Wu, Uncertainty quantification in intelligent-based electrical resistivity tomography image reconstruction with Monte Carlo dropout strategy, *IEEE Trans. Geosci. Rem. Sens.* 61 (2023) 1–16.