ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Fault Diagnosis of Dynamic Chemical Processes Based on Improved Residual Network Combined with a Gated Recurrent Unit

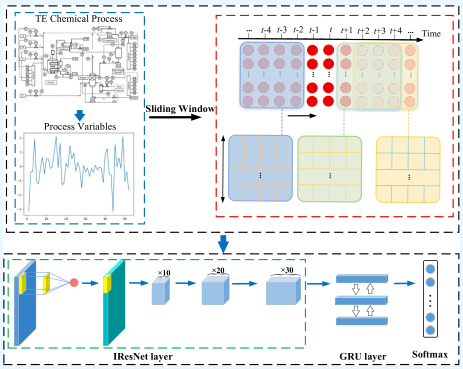
Shiqian HAN, Pingping WANG,* Cheng ZHANG, and Jun WANG

Cite This: *ACS Omega* 2025, 10, 8859−8869

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Aiming at the challenge of distinguishing the contributions of variables in dynamic chemical process data, this paper proposes a novel fault diagnosis method based on the IResNet-GRU model. First, we utilize principal component analysis to compute a correlation matrix, which serves as the input for an attention module. This approach enables the evaluation of feature contributions to predictions, thereby identifying the root-cause variables responsible for faults. Concurrently, we enhance the residual network (ResNet) with the attention module to assign weights to the extracted features. The improved ResNet (IResNet) can differentiate the significance of the monitored variables. Second, we augment the raw data into two-dimensional data using sliding window technology, capturing spatial and temporal data features. Finally, a gated recurrent unit is integrated to extract dynamic features from the augmented two-dimensional data effectively. The proposed method is validated using the Tennessee−Eastman chemical process. The diagnosis results demonstrate that the proposed method outperforms conventional methods.



1. INTRODUCTION

With the advancement of modern industrial technology, chemical processes have become increasingly complex and highly automated. Once an accident occurs, it can adversely affect production safety, production efficiency and the quality of products.[1] Therefore, conducting in-depth research on fault diagnosis methods and promptly diagnosing faults in the chemical production process are extremely important.

Fault diagnosis methods are typically classified into three categories: knowledge-based method,[2] model-based method,[3] and data-driven method,[4] as summarized by Venkatasubramanian et al. However, due to challenges in modeling and the heavy reliance on expert experience, both knowledge-based and model-based methods have limited applicability. In contrast, data-driven methods, which analyze and diagnose based on actual observed data, have gradually become a research focus.[5] For this reason, multivariate statistical process monitoring methods such as principal component analysis (PCA)[6,7] and independent component analysis (ICA)[8] have been widely applied in industrial process fault diagnosis. These methods achieve fault diagnosis by converting the original data from a high-dimensional space to a low-dimensional feature space. However, they are only suitable for identifying linear features, resulting in lower fault diagnosis rates for nonlinear and non-Gaussian complex industrial process data. In recent years, the rapid development of machine learning has positioned deep learning as an effective approach to overcoming these limitations. The essence of deep learning lies in networks autonomously learning abstract features from raw data at multiple levels. Traditional fault diagnosis methods in deep learning include autoencoders (AE),[9,10] convolutional neural networks (CNN),[11,12] and recurrent neural networks (RNN).[13,14] Yu et al. designed a one-dimensional CNN based on wavelet transform with multiple channels, applying it to the batch fermentation process of penicillin and the Tennessee−Eastman (TE) chemical process. The results indicate that this method is capable of learning signal features in high dimensions.[15] Yu and Zhao proposed a broad CNN with incremental learning capability, which can self-update when encountering new faults, including novel abnormal samples and fault categories.[16] Furthermore, residual network (ResNet) as a specialized variant of CNN, effectively addresses the issue of network degradation. Pan et al. proposed an end-to-end rolling bearing fault diagnosis algorithm based on an improved ResNet, which replaces the fully connected layers of traditional ResNet with global average pooling technique, effectively reducing the network parameters.[17] The hybrid attention improved residual network (HA-ResNet) proposed by Zhang

Received: April 19, 2024
Revised: October 24, 2024
Accepted: October 29, 2024
Published: February 24, 2025

et al. is designed for diagnosing faults in wind turbine gearboxes, with a focus on highlighting the fundamental frequency bands of wavelet coefficients and capturing faulty features within convolutional channels.[18] However, all the above methods only statically learn data features and rely on the quality of feature extraction, neglecting the dynamic representation of process data. In fact, the data collected in chemical processes is usually dynamic, which means that the occurrence of current failures depends on the previous state of the system. Taking this factor into consideration, recurrent neural network (RNN) fully considers the correlation between temporal data, making it more suitable for dynamic process fault diagnosis.[19] The gated recurrent unit (GRU) is a variant of RNN that automatically retains relevant information in dynamic sequential data while eliminating extraneous details.[20] Yuan et al. proposed a three-stage fault diagnosis method based on GRU networks, which was applied to the TE chemical process and the oxidation process of *para*-xylene (PX). The results indicate that this method performs well even with a limited historical data.[21] Zhang et al. combined the GRU with an enhanced deep convolutional neural network (EDCNN) for fault diagnosis in both the TE process and acid gas absorption process. The effectiveness of this method was validated in real-time industrial processes.[22] To effectively model multivariate time sequences with heterogeneous sampling intervals and missing values, Yuan et al. proposed an attention-based interval-aided network (AIA-Net).[23] This network incorporates mechanisms such as attention-based time-aware dynamic imputation and interval-aided time-aware networks to enhance temporal information processing. In addition, Yuan et al. proposed a multiscale attention-based CNN (MSACNN), which employed a channel-wise attention mechanism to weigh the significance of local spatiotemporal features at different scales.[24] Furthermore, to accurately predict critical quality variables, a novel soft sensor modeling method based on VCA-CNN[25] was proposed by Yuan et al. This method demonstrated its effectiveness in the hydrocracking and debutanizer column processes by leveraging time delay techniques and correlation analyses.However, the coupling relationship among each variable in the chemical production process results in differential contributions of information to the results of fault diagnosis. the aforementioned methods of processing dynamic data only connect features in channel dimension, lacking a suitable mechanism to reflect the correlation and importance of features across different channels. It leads to users lacking information about the contribution of variables to predictions, leaving them unaware of how to prevent the problem from recurring without analyzing potential root-cause features.

From the above analysis, this article proposes a novel fault diagnosis model, namely, IResNet-GRU, which combines the superior feature extraction capability of ResNet and the effectiveness of GRU in handling sequential data. First, we utilize PCA to optimize the attention mechanism and construct an improved ResNet (IResNet) network to address the issue of unclear feature importance. It not only balances dynamic fault information across the channel dimension, but also reduces attention on redundant features and identifies the root-cause features. Then, we use a sliding window to augment the raw data into a two-dimensional grid-like data in order to enhance the temporal correlation information on the data. In this way, the augmented data can provide abundant spatial and temporal feature information for the network. Finally, the dynamic information is further extracted in the temporal dimension using GRU. The superiority of the proposed method is ultimately verified through experiments on the TE chemical process.

In summary, the key contributions of this article are as follows.

1) Based on the two-dimensional grid data obtained by sliding window processing, a dynamic chemical process fault diagnosis model combining improved residual network and GRU is proposed, which can effectively capture the dynamic spatial characteristics of the system.
2) The variable contribution matrix is computed through PCA to assess feature importance and then used as input for the attention module. This process enables the model to adaptively assign weights to the extracted features.
3) The constructed model is applied to the fault diagnosis of Tennessee−Eastman process, and the performance comparison with other deep learning methods proves the advantages of the proposed method.

2. RELEVANT THEORY

2.1. Principal Component Analysis. The principal component analysis (PCA) method transforms the raw data into a new coordinate system through linear transformation.[26] This method uses dimensionality reduction to retain the most influential features on data variance, reflecting its root-cause variables. Assuming $\boldsymbol{x} = [x_i(k)]_{n\times m}(k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$ represents a data matrix with n samples and m variables. First, use eq 1 to standardize $\boldsymbol{x}$, where μ and σ are the mean and standard deviation of x_i. The matrix $\boldsymbol{X}$ is defined as the result of standardized $\boldsymbol{x}$.

$$x_i'(k) = \frac{x_i(k) - \mu}{\sigma} \tag{1}$$

Next, calculate the covariance matrix $\boldsymbol{C}$ of matrix $\boldsymbol{X}$ using eq 2.

$$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X} \tag{2}$$

Assuming that the projection of $\boldsymbol{X}$ in a one-dimensional space is $\boldsymbol{Y}$, the projection relationship can be expressed as $\boldsymbol{Y} = \boldsymbol{Xu}$, where $\boldsymbol{u}$ represents the projection vector. The objective of PCA is to find the projection vector $\boldsymbol{u}$ that maximizes the variance of $\boldsymbol{Y}$. To ensure the uniqueness of the solution for $\boldsymbol{u}$, it is necessary to introduce constraints $\mathbf{u}^T\mathbf{u} = 1$. The objective function of PCA is described as follows:

$$\begin{aligned} &\max E(\boldsymbol{Y} - E(\boldsymbol{Y}))^2 \\ &= \max \frac{1}{n-1} \times \boldsymbol{u}^T \times \boldsymbol{X}^T \times \boldsymbol{X} \times \boldsymbol{u} \\ &= \max \boldsymbol{u}^T \times \boldsymbol{C} \times \boldsymbol{u} \end{aligned} \tag{3}$$

Construct the Lagrangian function as shown in eq 4.

$$L(\boldsymbol{u}, \lambda) = \boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u} - \lambda(\boldsymbol{u}^T\boldsymbol{u} - 1) \tag{4}$$

Then, the partial derivatives of $\boldsymbol{u}$ and λ can be obtained using as follows:

$$\frac{\partial L(\boldsymbol{u}, \lambda)}{\partial \boldsymbol{u}} = 2\boldsymbol{C}\boldsymbol{u} - 2\lambda\boldsymbol{u} = 0 \tag{5}$$

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

$$\frac{\partial L(\boldsymbol{u}, \lambda)}{\partial \lambda} = \boldsymbol{u}^T\boldsymbol{u} - 1 = 0 \tag{6}$$

According to the partial derivative equation, $\boldsymbol{Cu} = \lambda\boldsymbol{u}$ can be obtained, where $\boldsymbol{u}$ is the eigenvector of covariance matrix $\boldsymbol{C}$, while λ is the eigenvalue of $\boldsymbol{C}$. Therefore, the objective function (eq 3) can be expressed as follows.

$$\max \boldsymbol{u}^T\boldsymbol{Cu} = \max \boldsymbol{u}^T\lambda\boldsymbol{u} = \max \lambda\boldsymbol{u}^T\boldsymbol{u} = \max \lambda \tag{7}$$

Consequently, the eigenvector $\boldsymbol{u}$ corresponding to the largest eigenvalues$\lambda$are the directions of principal component projection. In brief, projecting $\boldsymbol{X}$ toward the vector $\boldsymbol{u}$ yields the contribution scores for the most significant features.

2.2. ResNet. Residual network (ResNet) is a powerful architecture of convolutional neural networks that exhibits outstanding accuracy and convergence. ResNet can effectively address the problem of network degradation.[27] The main components of ResNet are as follows.

2.2.1. Convolutional Layers. The input of convolutional layers is a three-dimensional data with dimensions $H \times W \times C$, where H represents height, W represents width and C represents the number of channels. The equation for the convolution layers is defined as eq 8.

$$\boldsymbol{y}^i = f\left(\sum \boldsymbol{w}^i \times \boldsymbol{x}^{i-1} + \boldsymbol{b}^i\right) \tag{8}$$

where i is the layer index of network and $\boldsymbol{x}$ and $\boldsymbol{y}$ are the input and output of feature maps, while $\boldsymbol{w}$ is the weight matrix, $\boldsymbol{b}$ is the bias term, and $f(\cdot)$ is the activation function.

2.2.2. Shortcut Connections. Shortcut connections are the core part of the ResNet model. Also, it is key to the superior performance of ResNet over the traditional CNN model.[28] ResNet introduces shortcut connections in the CNN framework, which no longer fit the underlying network and instead adjust the corrective residuals based on original input mapping. If the input mapping is already optimal, the residual module fits a zero mapping and directly skips multiple layers of network propagation to reach the output. This identity mapping is shown as eq 9. Also, the structure of the residual module is shown in Figure 1.

$$\boldsymbol{y} = F(\boldsymbol{x}, \{\boldsymbol{W}_i\}) + \boldsymbol{x} \tag{9}$$

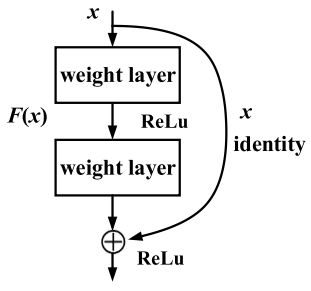


Figure 1. Structure of the residual module.

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the input and output of the residual blocks, $\boldsymbol{W}_i$ denotes the weight matrix, and $F(\cdot)$ represents the mapping of residual in the fitting process. Assuming a stack of multiple residual blocks, the information is propagated forward from residual block i to the residual block j, as shown in eq 10.

$$\boldsymbol{x}_j^{\text{out}} = \boldsymbol{x}_i^{\text{in}} + \sum_{k=i}^{j} F(\boldsymbol{x}_k^{\text{in}}, \{\boldsymbol{W}_k\}) \tag{10}$$

where $\boldsymbol{x}_i^{in}$ and $\boldsymbol{x}_k^{in}$ are the inputs of the residual block i and j,$\boldsymbol{x}_j^{out}$ is the output of the residual block j, $\boldsymbol{W}_k$ is the weight matrix of the residual block k, and F is the residual mapping. In backpropagation, the gradient formula for optimizing the network in terms of loss ε is defined as eq 11.

$$\frac{\partial \varepsilon}{\partial \boldsymbol{x}_i^{\text{in}}} = \frac{\partial \varepsilon}{\partial \boldsymbol{x}_j^{\text{out}}}\frac{\partial \boldsymbol{x}_j^{\text{out}}}{\partial \boldsymbol{x}_i^{\text{in}}} = \frac{\partial \varepsilon}{\partial \boldsymbol{x}_j^{\text{out}}}\left[1 + \frac{\partial}{\partial \boldsymbol{x}_i^{\text{in}}}\sum_{k=i}^{j} F(\boldsymbol{x}_k^{\text{in}}, \{\boldsymbol{W}_k\})\right] \tag{11}$$

where $1 + \frac{\partial}{\partial \boldsymbol{x}_i^{\text{in}}}\sum_{k=i}^{j} F(\boldsymbol{x}_k^{\text{in}}, \{\boldsymbol{W}_k\})$ is not less than 1 to prevent the occurrence of gradient vanishing.

3. PROPOSED METHOD

The proposed fault diagnosis method for dynamic chemical processes, known as IResNet-GRU, consists of two parts: sliding window and IResNet-GRU model. As shown in Figure 2, first, the original one-dimensional single-point data is transformed into two-dimensional grid-like data through sliding window processing, which serves as the input for the model. Then, the contribution of variables is distinguished in the IResNet layers to fully extract features from the root-cause variables. Finally, the feature maps generated by the IResNet layers are inputted into the GRU layers to further extract dynamic temporal information. The feature information is combined with temporal information through a fully connected layer, followed by input into a Softmax classifier for classification purposes. The proposed method finally outputs the diagnostic probabilities for each predefined fault category.

3.1. Sliding Window. In chemical production processes, the dynamic characteristics of data make it challenging to capture features using a single global diagnostic model. For this reason, we introduce sliding window technology. The raw single-point one-dimensional data is augmented in the temporal dimension to provide sufficient dynamic temporal features for the network. A schematic diagram of sliding window processing is shown in Figure 3. Generally, for a given multivariate time series $\boldsymbol{X} = \{\boldsymbol{x}_t | t = 1,2,3\cdots n\}$, where $\boldsymbol{x}_t$ denotes the process monitoring vector at time t and n denotes the length of $\boldsymbol{x}_t$. Let $\mathbf{Y} = \{\boldsymbol{y}_t | t = 1,2,3\cdots n\}$be the label of $\boldsymbol{X}$. As shown in Figure 3, a sliding window of length m and width $d(0 < d \leq n)$ partitions the sequence $\boldsymbol{X}$ into n subsequences $\boldsymbol{s}_n$ $(n = 1,2,3\cdots N)$along the temporal dimension. Then we establish a feature map for each subsequence and let $\boldsymbol{Y}_d = (\boldsymbol{y}_d | d = 1,2,3\cdots N)$ be the label of $\boldsymbol{s}_n$. The sliding window moves with an integer step size of λ $(0 < \lambda \leq n - d)$. Meanwhile, the dimensions of feature information and temporal information can be adjusted by varying the sizes of m and d. In summary, the original one-dimensional single-point data contains only feature information. After sliding window processing, the data is transformed into two-dimensional grid-type data that integrates both feature and temporal information. In other words, these two-dimensional samples can be interpreted as the current system state is influenced by the previous d moments system states. The proposed fault diagnosis model is trained using these samples, enabling simultaneous learning of both feature and temporal information. This approach significantly enhances the performance of fault diagnosis.

3.2. IResNet-GRU Model. In this section, we design an improved ResNet (IResNet) to address the challenge of undifferentiated feature importance. To be specific, the ResNet

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

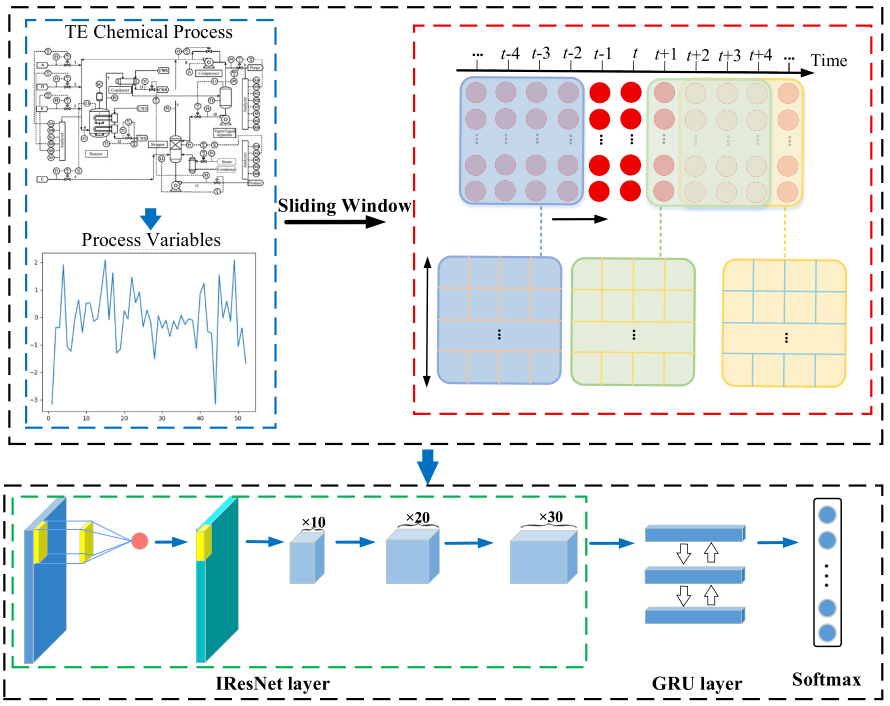


Figure 2. Framework of the proposed IResNet-GRU-based fault diagnosis method.

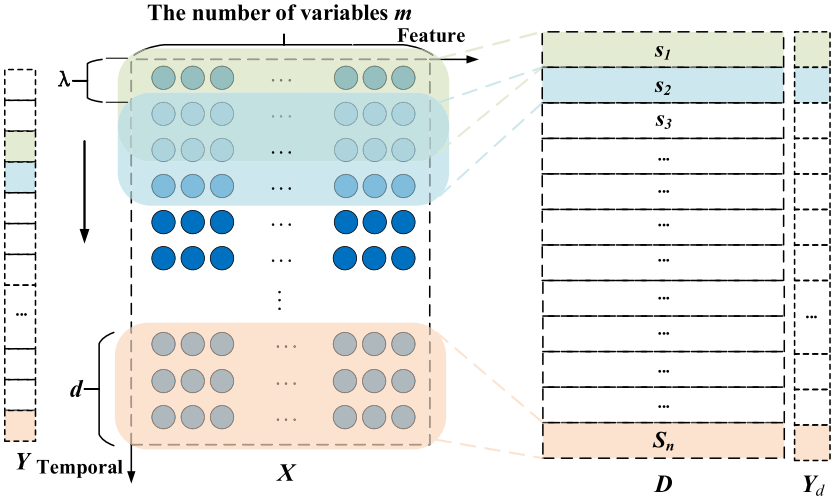


Figure 3. Schematic diagram of sliding window processing.

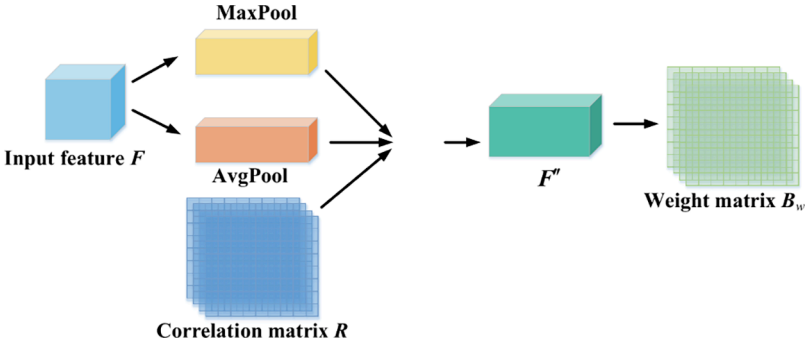


Figure 4. Operation of the improved attention module.

is optimized by utilizing the improved attention module, which adjusts the input feature maps by calculating attention weights. This work allows the optimized IResNet layers to allocate more attention to prominent features. Then, the GRU is incorporated into the model framework in order to further capture the dynamic features of data. In the subsequent

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

sections, a comprehensive description of the proposed model will be provided.

3.2.1. IResNet Layers. The IResNet layers consist of two parts, namely, the improved attention module and the ResNet module. The operation of the improved attention module is shown in Figure 4. The input consists of a feature map $\boldsymbol{F}$ and a correlation matrix $\boldsymbol{R}$ representing the attention region, with dimensions $B \times C \times W \times H$ and $B \times W \times H$, respectively, where B denotes the batch size, while C, W, and H represent the number of channels, width, and height. In particular, The correlation matrix $\boldsymbol{R}$ is obtained by applying the PCA algorithm to calculate. The matrix $\boldsymbol{R}$ contains information regarding the relevance of each time step to fault occurrences. The specific calculation process is shown in Section 4.

To illustrate the improved attention module clearly, we initially flatten the feature map $\boldsymbol{F}$ in spatial dimensions to a size of $B \times C \times WH$. For each channel C, adaptive average pooling and adaptive max pooling are used separately to compress a one-dimensional vector of length WH into a scalar value. Obviously, the most prominent features are extracted by adaptive max pooling. A vector of shape $B \times 1 \times C$ is obtained. Next, the vector is activated using the *sigmoid* function to obtain two vectors $\boldsymbol{F}'_{\mathrm{AVG}}$ and $\boldsymbol{F}'_{\mathrm{MAX}}$ of shape $B \times WH \times C$ as shown in eq 12. Subsequently, $\boldsymbol{F}'_{\mathrm{AVG}}$ and $\boldsymbol{F}'_{\mathrm{MAX}}$ are reshaped into vectors with dimensions $B \times W \times H$ and added to the corresponding matrix $\boldsymbol{R}$. In this way, we obtain a fusion matrix $\boldsymbol{F}''$ of shape of $B \times W \times H$. Finally, we use the *sigmoid* activation function to obtain an attention weight matrix $\boldsymbol{B}_w$ with a shape of $B \times WH$. It can be observed that the weight matrix $\boldsymbol{B}_w$ is capable of guiding the network to focus on learning features with greater contribution. Finally, as shown in eq 14, the feature map is multiplied by the attention weight matrix $\boldsymbol{B}_w$ for each channel C to obtain a weighted feature map $\boldsymbol{F}_{\mathrm{output}}$.

This section constructs the IResNet layers by integrating the improved attention module into ResNet. And the enhancement improves the learning capability of IResNet layers for identifying root-cause variables, while minimizing interference from redundant data in network classification tasks, thereby enhancing diagnostic efficiency. In conclusion, the IResNet layers assign corresponding weights to different spatial positions, effectively addressing issues of poor distinguishable feature importance.

$$\boldsymbol{F}'_{\mathrm{AVG}} = \sigma(\mathrm{AvgPool}(\boldsymbol{F})) \tag{12}$$

$$\boldsymbol{F}'_{\mathrm{max}} = \sigma(\mathrm{maxPool}(\boldsymbol{F}) \tag{13}$$

$$\boldsymbol{F}_{\mathrm{output}} = \boldsymbol{F} \odot \boldsymbol{B}_w \tag{14}$$

3.2.2. GRU Layers. Gated recurrent unit (GRU) is a variant of recurrent neural networks (RNN) designed for processing sequential data. Similar to the long short-term memory (LSTM) model, GRU includes a dedicated memory unit to handle long-term dependencies. It utilizes gate mechanisms to control the flow of information, thereby improving the model accuracy and performance.[29] Compared to LSTM, GRU effectively captures dynamic features in time series data with fewer parameters.[30] Given the significant temporal information inherent in chemical process data faults, GRU is well-suited for managing sequential data. Therefore, the GRU model is selected for this study to capture dynamic characteristics. The GRU consists of two parts: the update gate and the reset gate. The forward propagation equations of GRU network are defined as follows:

$$\boldsymbol{R}_t = \sigma(\boldsymbol{W}_r[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{15}$$

$$\boldsymbol{Z}_t = \sigma(\boldsymbol{W}_z[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{16}$$

$$\bar{\boldsymbol{h}}_t = \tanh(\boldsymbol{W}_{\bar{h}}[\boldsymbol{R}_t \times \boldsymbol{h}_{t-1}, \boldsymbol{x}_t]) \tag{17}$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{Z}_t) \times \boldsymbol{h}_{t-1} + \boldsymbol{Z}_t \times \boldsymbol{h}_t \tag{18}$$

where $\boldsymbol{x}_t$ is the input at current time t, while $\boldsymbol{h}_t$ and $\boldsymbol{h}_{t\text{-}1}$ respectively represent the hidden state at the current and previous moments, $\boldsymbol{R}_t$ denotes the reset gate, and $\boldsymbol{Z}_t$ represents the update gate. Meanwhile, $\bar{\boldsymbol{h}}_t$ represents the output of candidate hidden states and $\boldsymbol{W}_r$, $\boldsymbol{W}_z$, and $\boldsymbol{W}_{\bar{h}}$ represent the weight matrices of the reset gate, update gate, and candidate output set. The structure of GRU is shown in Figure 5. In brief,

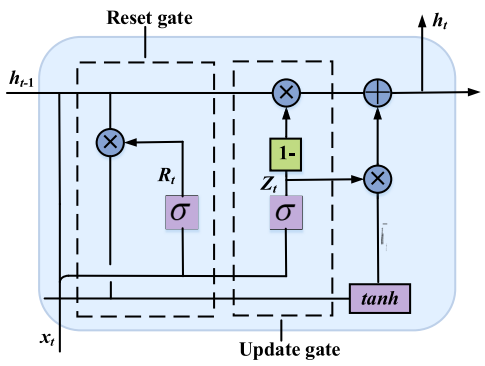


Figure 5. Structure of GRU.

the GRU uses the last hidden state and the current input to obtain two gating states. Then it calculates the candidate set data using the reset gate, and finally, it can obtain a new state through the update gate.

3.3. Method Framework. The production process data of the chemical industry are dynamic and undifferentiated in importance of variables. Therefore, this article constructs the IResNet-GRU method, introducing the sliding window to obtain augmented data dynamically and using the improved attention mechanism to optimize ResNet. The proposed method utilizes the PCA algorithm to calculate the correlation matrix of variables to reduce the dimension and inputs the correlation matrix into the improved attention module. The model can assign greater weights to critical channel fault features by reducing attention to redundant features, thereby improving the accuracy of the model. More importantly, it can quickly identify the root-cause variables that lead to faults, allowing users to address them promptly. Figure 6 illustrates the method framework, which mainly consists of two stages: offline modeling and online diagnosis.

Offline modeling stage:

Step 1 Historical data of TE chemical process was collected under normal operating conditions and fault conditions.

Step 2 The raw data was preprocessed, and the correlation matrix $\boldsymbol{R}$ was calculated based on the historical data.

Step 3 The network architecture was defined, including the number of layers, size of convolutional kernels, dimensions of input data, and activation functions.

Step 4 Training parameters such as learning rate, batch size, epoch, the width of sliding window, and optimizer were set.

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

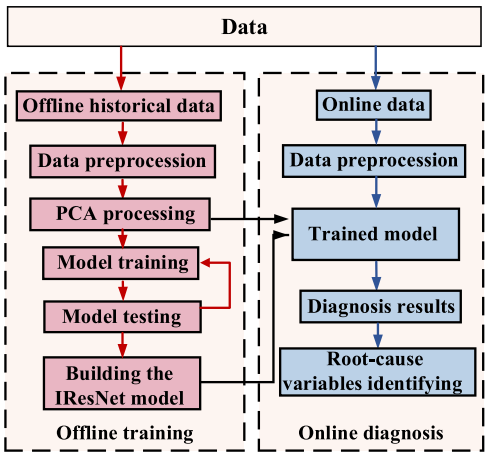


Figure 6. Method framework.

Step 5 The model was trained by adjusting training parameters.

Online diagnosis stage:

Step 1 Process data was collected online.

Step 2 The online data and the correlation matrix $\boldsymbol{R}$ was inputted into the trained IResNet-GRU model.

Step 3 The fault data was diagnosed, and the root-cause variables for these faults was identified based on the matrix $\boldsymbol{R}$.

4. SIMULATION EXPERIMENT ANALYSIS

4.1. Data Set Introduction. The data of Tennessee–Eastman (TE) chemical process exhibits characteristics such as temporal variation, strong coupling, non-Gaussian distribution, and nonlinearity. The TE data set consists of 22 distinct simulation runs, including both normal production processes and 21 fault processes induced by deliberate disturbances. Accordingly, the TE chemical process can be observed to obtain a vector that reflects the production situation at time t as follows:

$$\boldsymbol{x}_t = [\mathrm{XMV}(1), \cdots, \mathrm{XMV}(11), \mathrm{XMEAS}(1), \mathrm{XMEAS}(2), \cdots, \mathrm{XMEAS}(41)]^T \quad (19)$$

where XMV is manipulated variables and XMEAS is measured variables. The sampling frequency for both XMV and XMEAS is 20 times per hour. The detailed description of all faults is shown in Table 1. The data of the TE chemical process is divided into training and testing sets, with each training set consisting of 480 samples and each testing set consisting of 960 samples. There are a total of 22 system states, resulting in 10,560 original training samples and 21,120 original testing samples.

Table 1. Process Faults for the TE Process Simulator[34]

fault number	fault desription	type
1	A/C feed ratio, B composition constant	step
2	B composition, A/C ratio constant	step
3	D feed temperature	step
4	reactor cooling water inlet temperature	step
5	condenser cooling water inlet temperature	step
6	A feed loss	step
7	C header pressure loss-reduced availability	step
8	A, B, C feed composition	random variation
9	D feed temperature	random variation
10	C feed temperature	random variation
11	reactor cooling water inlet temperature	random variation
12	condenser cooling water inlet temperature	random variation
13	reaction kinetics	slow drift
14	reactor cooling water valve	sticking
15	condenser cooling water valve	sticking
16–20	unknown	unknown
21	the valve for stream 4 was fixed at the steady-state position	constant position

4.2. Feature Selection. Since the intricate interdependencies among variables and the pronounced high-dimensional nonlinearity in chemical processes, the results of fault diagnosis are influenced differentially by distinct variables. The primary objective of feature selection is to achieve a precise reduction in dimensionality, thereby effectively enhancing the performance of fault diagnosis. Therefore, we use PCA to perform feature selection on the original high-dimensional nonlinear data. First, calculate the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n$ corresponding to the covariance matrix $\boldsymbol{C}$ in eq 7. Then, the number of principal components is able to be determined using the cumulative percent variance (CPV) method. The CPV threshold in this article is set to 0.9,[31−33] retaining the variable information that contributes most to the occurrence of faults. Finally, create the matrix $\boldsymbol{R}$ to store information about these contributions. The correlation matrix $\boldsymbol{R}$ in Figure 7 visually represents the variables' contributions to the prediction results in a more comprehensible way. The role of $\boldsymbol{R}$ is to guide network learning by marking an attention region in the attention module, thereby reducing dimensions and minimizing interference caused by redundant features while improving model training speed and accuracy. It should be noted that matrix $\boldsymbol{R}$ contains information about the root-cause variables that led to the occurrence of faults. $\boldsymbol{R}$ can efficiently and precisely locate root-cause variable factors that affect faults, thereby maximizing the reduction of losses caused by faults.

4.3. Sliding Window Processing. The training and testing sets are processed using sliding windows with a step size of 1 and a window length of 12. In this context, the number of variables $m = 52$, represents the number of variables in each sample.The sliding window results for the original training and testing sets are shown in Table 2. It is worth noting that the obtained two-dimensional grid data incorporates information about the features as well as dynamic temporal information due to the implemention of the sliding window technique. Consequently, the model is provided with suffcient diagnostic information by including more detailed data.

4.4. Experiment Analysis. The detailed settings of each layer of the IResNet-GRU model are shown in Table 3. For the hyperparameter settings, the batch size is 64, the number of epochs is 200, the learning rate is 0.001. Cross-entropy loss function is defined as follows:

$$L = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \quad (20)$$

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

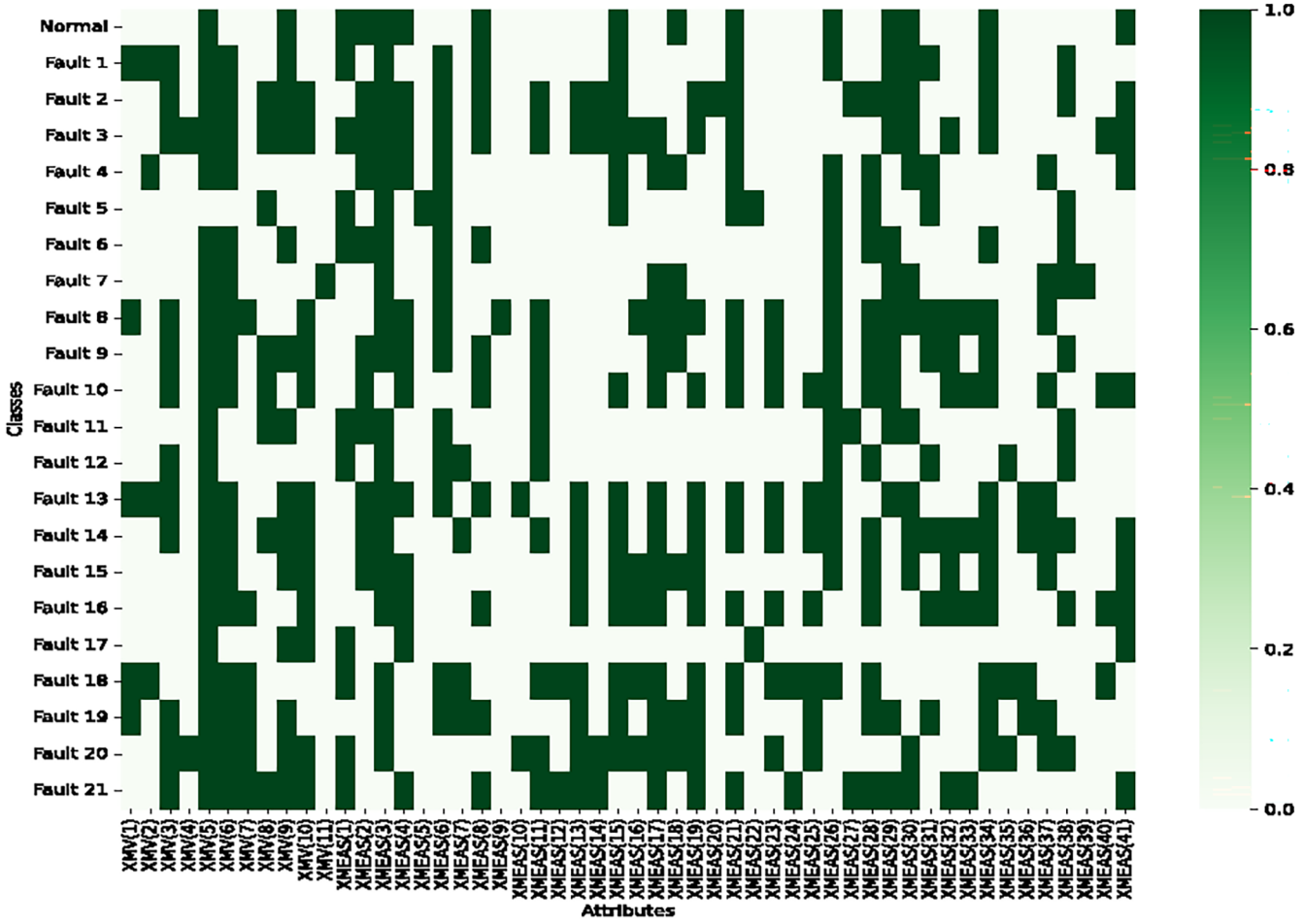


Figure 7. Correlation matrix R.

Table 2. Result of Sliding Window Processing

classes	data set	original length	number of samples after sliding window
normal or faults 1−21	training set	480	469
	testing set	960	949

Table 3. Detailed Settings of the IResNet-GRU Model

layer	name of layer	output shape	kernel size/kernel number/ stride/zero-padding
layer 1	IResNet_1	(none,10,52,52)	(3,3) /10/1 /1
layer 2	MaxPool2d	(none,10,26,26)	
layer 3	IResNet_2	(none,20,26,26)	(3,3) /20/1 /1
layer 4	MaxPool2d	(none,20,13,13)	
layer 5	IResNet_3	(none,30,13,13)	(3,3) /30/1 /1
layer 6	Dense_1	(none,1000)	
layer 7	GRU_1	(none,12,100)	
layer 8	Dense_2	(none,1000)	
layer 9	Dropout	(none,1000)	
layer 10	Dense_3	(none,22)	

where M represents the number of classes and y_{ic} is set to 1 if the true class of sample i is equal to c, otherwise, it is set to 0. Also, p_{ic} is the predicted probability that observation sample i belongs to category c. Additionally, Adam optimizer is employed for training. Meanwhile, we implemented L2 regularization on the GRU layers and incorporated dropout techniques in the fully connected layers to mitigate potential overfitting issues.

4.4.1. Evaluation Indexes. There are four evaluation metrics for classification problems: TP (true positive), FP (false positive), TN (true negative), and FN (false negative). In this context, P and N represent the judgment results, while T and F indicate the correctness of the judgment. These four categories of metrics can generate multiple metrics for evaluating the network, including the accuracy (Acc), precision (P), recall (R), and F1-score (F1). The definitions of these four indexes are provided in eqs 21−24.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{21}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{22}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{24}$$

4.4.2. Results and Discussion. In order to verify the effectiveness of the proposed method for fault diagnosis in chemical processes, comparative experiments were conducted using CNN, ResNet, IResNet, GRU and the proposed

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

Table 4. Precision and Recall of All Models

	CNN		ResNet		IResNet		GRU		IResNet-GRU	
classes	P	R	P	R	P	R	P	R	P	R
normal	0.615	0.423	0.874	0.839	0.901	0.894	0.775	0.642	0.858	0.939
fault 1	0.900	0.983	1.000	0.945	1.000	0.930	0.920	0.981	1.000	0.973
fault 2	1.000	0.956	1.000	0.997	0.968	0.987	1.000	0.948	0.986	0.983
fault 3	0.157	0.159	0.856	0.883	0.932	0.948	0.468	0.417	0.947	0.962
fault 4	0.884	0.990	1.000	0.993	0.905	0.907	0.861	0.896	0.958	0.968
fault 5	0.751	0.738	0.948	0.942	0.980	0.970	0.975	0.979	1.000	0.967
fault 6	1.000	1.000	0.829	0.895	0.915	0.898	1.000	0.985	0.923	0.983
fault 7	1.000	0.994	0.997	0.990	0.904	0.916	1.000	1.000	0.925	0.977
fault 8	0.620	0.571	0.751	0.894	0.687	0.638	0.789	0.640	0.963	0.970
fault 9	0.130	0.184	0.365	0.480	0.918	0.933	0.395	0.424	0.950	0.975
fault 10	0.524	0.667	0.522	0.431	0.599	0.540	0.797	0.757	0.957	0.964
fault 11	0.749	0.755	0.661	0.600	0.758	0.776	0.830	0.755	1.000	0.957
fault 12	0.618	0.730	0.501	0.890	0.991	0.953	0.825	0.858	0.991	0.961
fault 13	0.876	0.444	0.998	0.919	0.978	0.977	0.807	0.613	0.979	0.971
fault 14	0.998	0.998	0.664	0.532	0.784	0.690	1.000	0.992	0.982	0.967
fault 15	0.149	0.268	0.845	0.838	0.961	0.998	0.324	0.523	0.970	0.972
fault 16	0.618	0.574	0.997	0.978	0.998	0.997	0.580	0.569	1.000	0.963
fault 17	0.949	0.858	0.998	0.995	0.997	0.998	0.984	0.900	0.987	0.965
fault 18	0.944	0.736	0.903	0.919	0.996	0.998	0.997	0.741	1.000	0.978
fault 19	0.775	0.822	1.000	0.990	0.992	0.994	0.812	0.803	1.000	0.966
fault 20	0.808	0.590	0.981	0.993	0.995	0.996	0.764	0.774	1.000	1.000
fault 21	0.584	0.640	0.996	0.992	0.998	0.998	0.379	0.454	1.000	1.000
average	0.711	0.685	0.854	0.867	0.916	0.906	0.749	0.721	0.972	0.970

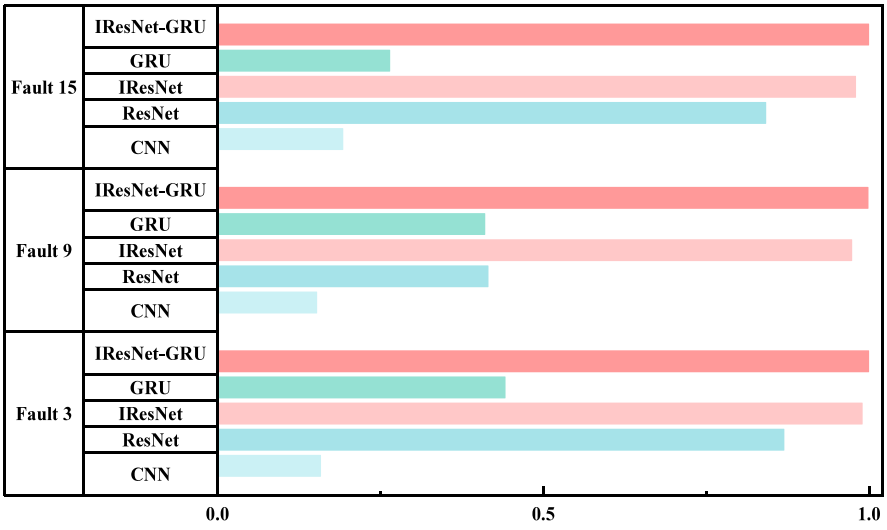


Figure 8. F1-score of five models for faults 3, 9, and 15.

IResNet-GRU model. Furthermore, both CNN and ResNet have the same number of network layers, consisting of five convolutional and pooling layers, as well as one Softmax classifier. The precision and recall rate of each model are presented in Table 4. It can be observed that the feature extraction capability of a single CNN model is insufficient for complex dynamic data in TE chemical process, resulting in its failure to achieve effective fault diagnosis. Compared to CNN, GRU has the ability to retain long-term historical information, allowing for better extraction of dynamic features from temporal data, thus resulting in slightly superior diagnostic performance. Meanwhile, the precision of ResNet is significantly improved compared to CNN since the shortcut connections, which enhance the learning process of the network and alleviate the problem of network degradation. We improve the attention module by the correlation matrix $\boldsymbol{R}$ to assign corresponding weights to spatial features, ultimately optimizing ResNet. The optimized IResNet can differentiate the importance of variables based on their contributions to prediction of faults. Therefore, it improves the ability to learn high-dimensional nonlinear chemical engineering data. However, the optimized IResNet model fails to extract temporal-related features from real chemical process data due to its dynamics. To fully capture the temporal information in TE data, this article introduces the sliding window technology to enhance the dynamism of the original data. Then, the IResNet module and GRU module are integrated to further improve the ability of the fused model to extract both spatial features

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

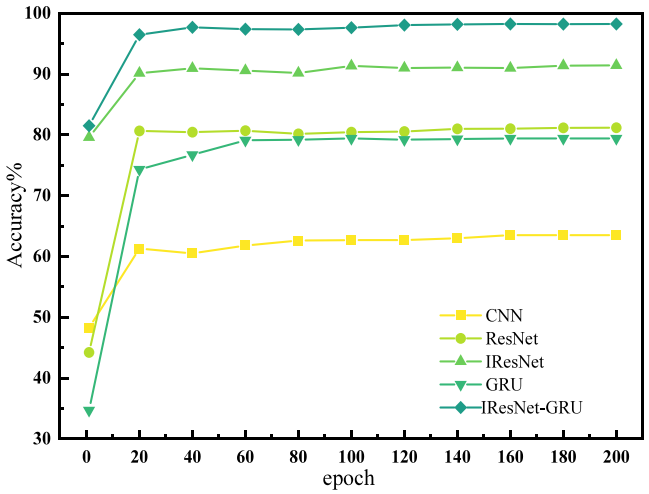


Figure 9. Accuracy of five models.

and dynamic temporal features. Obviously, the IResNet-GRU model proposed in this article achieves the most exceptional results compared to the other models.

As shown in Table 4, it can be seen that the CNN has poor diagnostic performance for faults 3, 9, and 15, indicating that a single convolutional layer is not efficient in learning features for such faults. The F1-score bar chart for these three types of faults is presented in Figure 8 to further illustrate. Due to the ability of GRU to sufficiently learn the temporal correlations between samples, it performs slightly better than CNN in diagnosing the three types of faults, resulting in an improved F1-score. The residual structure of ResNet enables the network to converge quickly and improves performance through learning identity mappings. Therefore, there is a significant improvement in the F1-score for faults 3 and 15, while the diagnostic performance remains poor for fault 9. We obtained IResNet model by using an improved attention mechanism to optimize ResNet. IResNet allows for the distinction of variable importance and increases the learning weight of key features in

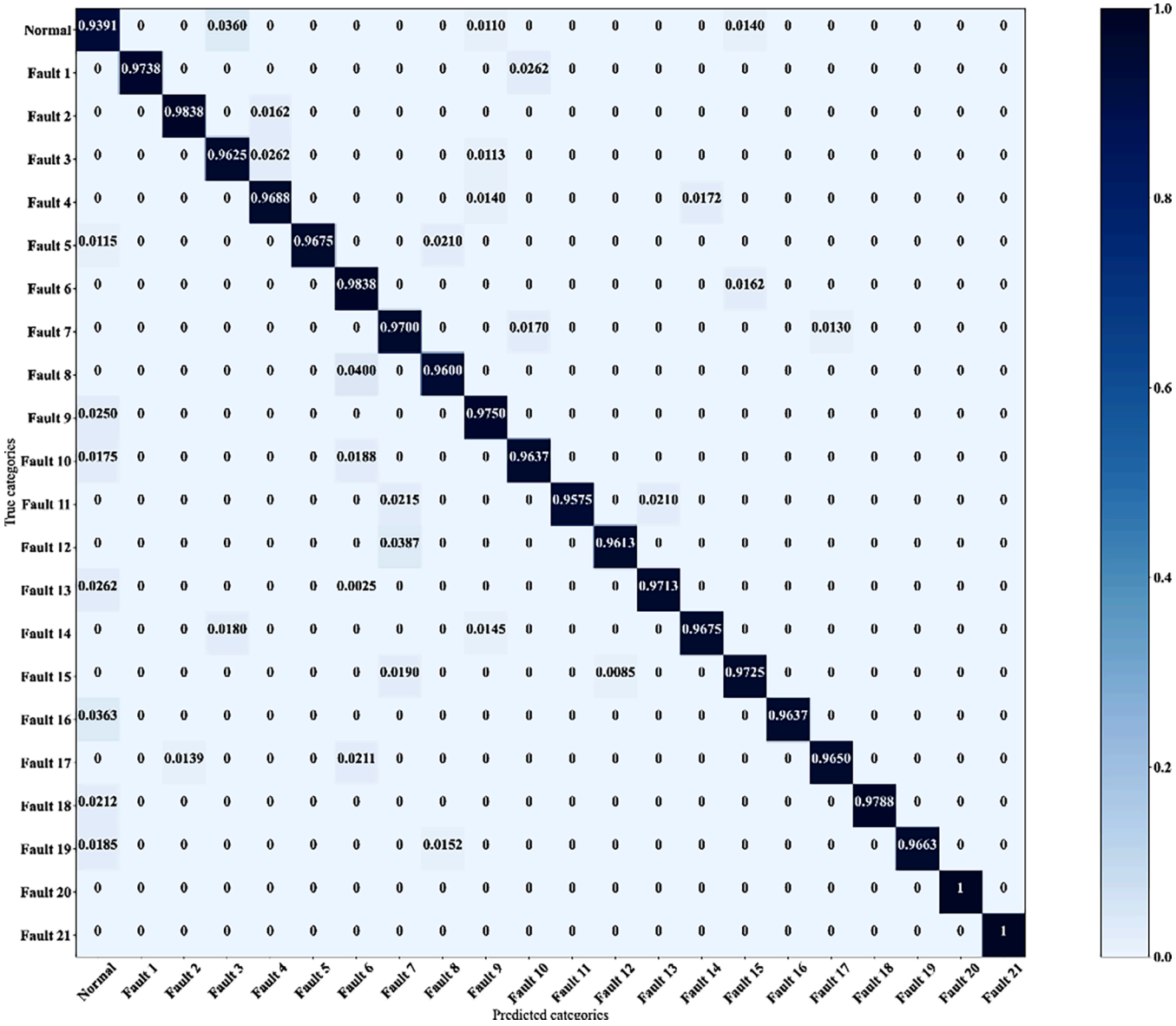


Figure 10. Confusion matrix of IResNet-GRU.

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

the model, resulting in better fault diagnosis results across three types of faults. In contrast, the proposed IResNet-GRU model synergistically integrates the spatial feature extraction capabilities of IResNet with the temporal feature extraction strengths of GRU. This combination results in the highest F1-score across the three challenging diagnostic fault types evaluated.

The accuracy comparison chart of all models on the same test set is shown in Figure 9. ResNet provides a direct path for gradients through shortcut connections, simplifying the learning task of the network. Compared to CNN, ResNet exhibits a significant improvement in convergence speed. The IResNet model employs an enhanced attention mechanism to differentiate the significance of variables while reducing their dimensions, thereby facilitating superior adaptation to intricate nonlinear functions. As a result, it performs well in terms of accuracy and convergence speed. On the other hand, the gate control mechanism of GRU can regulate the flow of information, capturing long-term dependencies in sequential data. It offers the advantages of fast convergence speed and strong model stability. Through comparison, the accuracy of the proposed model has consistently outperformed other networks since the first epoch. It indicates that the introduction of the sliding window and the improved attention mechanism contribute to enhancing the performance of the model. In conclusion, the IResNet-GRU model is suitable for fault diagnosis in dynamic nonlinear industrial processes.

In order to demonstrate the performance of the IResNet-GRU model proposed in this article for fault diagnosis of TE data, Figure 10 presents a confusion matrix that visually represents the diagnostic results for various types of faults. It can be observed that IResNet-GRU exhibits a high degree of accuracy in classifying faults. In fact, IResNet-GRU can quickly determine the root-cause variable based on the correlation matrix $\boldsymbol{R}$ during the online diagnosis stage. Therefore, the proposed method can effectively mitigate the losses resulting from faults.

5. CONCLUSIONS

This article introduces a novel fault diagnosis method based on IResNet-GRU for dynamic chemical process fault diagnosis. The method combines spatial and temporal features using a sliding window mechanism and enhances the attention module with PCA integrated into ResNet. In this way, high-dimensional nonlinear features can be more effectively extracted. Moreover, the network adaptively assigns weights to variables to highlight the contribution of root-cause variables. We also employ the GRU network to further extract dynamic features from chemical process data. In conclusion, the proposed method not only distinguishes the importance of variables and quickly identifies the variables causing faults but also captures the temporal correlation of samples. Furthermore, the proposed method can be embedded into the state monitoring systems of chemical processes to enable real-time analysis of potential faults and identification of the corresponding root-cause features.

AUTHOR INFORMATION

Corresponding Author

Pingping WANG − *College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang, Liaoning 110142, China; Key Laboratory for Chemical Process IndustryIntelligent Technology of Liaoning Province, Shenyang, Liaoning 110142, China;* orcid.org/0009-0009-3100-7313; Email: yzwping2024@163.com

Authors

Shiqian HAN − *College of Science, Shenyang University of Chemical Technology, Shenyang, Liaoning 110142, China; Key Laboratory for Chemical Process IndustryIntelligent Technology of Liaoning Province, Shenyang, Liaoning 110142, China*

Cheng ZHANG − *College of Science, Shenyang University of Chemical Technology, Shenyang, Liaoning 110142, China; Key Laboratory for Chemical Process IndustryIntelligent Technology of Liaoning Province, Shenyang, Liaoning 110142, China*

Jun WANG − *College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang, Liaoning 110142, China; Key Laboratory for Chemical Process IndustryIntelligent Technology of Liaoning Province, Shenyang, Liaoning 110142, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c03757

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Liaoning Province Nature Fund Project (no. 2022-MS-291); the National foreign expert project plan (G2022006008L); and the Scientific research project of Liaoning Province Education Department (LJKMZ20220781, LJKMZ20220783).

REFERENCES

(1) Qin, S. J. Process data analytics in the era of big data. *Aiche Journal* **2014**, *60* (9), 3092−3100.

(2) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Comput. Chem. Eng.* **2003**, *27* (3), 313−326.

(3) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27* (3), 293−311.

(4) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N.; Yin, K. A review of process fault detection and diagnosis: Part III: Process history based methods. *Comput. Chem. Eng.* **2003**, *27* (3), 327−346.

(5) Dai, X. W.; Gao, Z. W. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics* **2013**, *9* (4), 2226−2238.

(6) Li, W.; Peng, M. J.; Wang, Q. Z. Fault identification in PCA method during sensor condition monitoring in a nuclear power plant. *Annals of Nuclear Energy* **2018**, *121*, 135−145.

(7) Zhang, K.; Su, J. P.; Sun, S. A.; et al. Compressor fault diagnosis system based on PCA-PSO-LSSVM algorithm. *Sci. Prog.* **2021**, *104* (3), 1−16.

(8) Yu, G. Fault feature extraction using independent component analysis with reference and its application on fault diagnosis of rotating machinery. *Neural Computing and Applications* **2015**, *26*, 187−198.

(9) Lu, C.; Wang, Z. Y.; Qin, W. L.; et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing* **2017**, *130*, 377−388.

(10) Ma, J. P.; Li, C. W.; Zhang, G. Z. Rolling bearing fault diagnosis based on deep learning and autoencoder information fusion. *Symmetry* **2022**, *14* (1), 13.

(11) Choudhary, A.; Mishra, R. K.; Fatima, S.; et al. Multi-input CNN based vibro-acoustic fusion for accurate fault diagnosis of

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869

induction motor. *Engineering Applications of Artificial Intelligence* **2023**, *120*, No. 105872.
(12) Jin, T. T.; Yan, C. L.; Chen, C. H.; et al. Light neural network with fewer parametersbased on CNN for fault diagnosis of rotating machinery. *Measurement* **2021**, *181* (3), No. 109639.
(13) Zhang, S. Y.; Qiu, T. Semi-supervised LSTM ladder autoencoder for chemical process fault diagnosis and localization. *Chem. Eng. Sci.* **2022**, *251*, No. 117467.
(14) Khorram, A.; Khalooei, M.; Rezghi, M. End-to-end CNN plus LSTM deep learning approach for bearing fault diagnosis. *Applied Intelligence* **2021**, *51*, 736−751.
(15) Yu, J. B.; Zhang, C.; Wang, S. Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis ofindustrial processes. *Neural Computing and Applications.* **2021**, *33* (8), 3085−3104.
(16) Yu, W. K.; Zhao, C. H. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE Transactions on Industrial Electronics* **2020**, *67* (6), 5081−5091.
(17) Hao, X. Y.; Zheng, Y.; Lu, L.; et al. Research on intelligent fault diagnosis of rolling bearing based on improved deep residual network. *Applied Sciences* **2021**, *11* (22), 10889.
(18) Zhang, K.; Tang, B. P.; Deng, L.; et al. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement* **2021**, *179* (10), No. 109491.
(19) Chen, H. H.; Cen, J.; Yang, Z. H.; et al. Fault diagnosis of the dynamic chemical process based on the optimized CNN-LSTM network. ACS. *Omega* **2022**, *7* (38), 34389−34400.
(20) Ravanelli, M.; Brakel, P.; Omologo, M.; et al. Light gated recurrent units for speech recognition. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2018**, *2* (2), 92−102.
(21) Yuan, J.; Tian, Y. An intelligent fault diagnosis method using GRU neural network towards sequential data in dynamic processes. *Processes* **2019**, *7* (3), 152.
(22) Zhang, J. X.; Zhang, M.; Feng, Z. M.; et al. Gated recurrent unit-enhanced deep convolutional neural network for real-time industrial process fault diagnosis. *Process Safety and Environmental Protection* **2023**, *175*, 129−149.
(23) Yuan, X. F.; Xu, N.; Ye, L. J.; et al. Attention-based interval aided networks for data modeling of heterogeneous sampling sequences with missing values in process industry. *IEEE Transactions on Industrial Informatics* **2024**, *20* (4), 5253−5262.
(24) Yuan, X. F.; Huang, L. F.; Ye, L. J.; et al. Quality prediction modeling for industrial processes using multiscale attention-based convolutional neural network. IEEE Transactions on. *Cybernetics* **2024**, *54* (5), 2696−2707.
(25) Yuan, X. F.; Wang, Y. C.; Wang, C.; et al. Variable correlation analysis-based convolutional neural network for far topological feature extraction and industrial predictive modeling. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1−10.
(26) Zhang, C.; Guo, Q. X.; Li, Y. Fault detection in the tennessee eastman benchmark process using principal component difference based on k-nearest neighbors. IEEE. *Access* **2020**, *8*, 49999−50009.
(27) Yan, X. C.; Zhang, Y.; Jin, Q. B. Chemical process fault diagnosis based on improved ResNet fusing CBAM and SPP. IEEE. *Access* **2023**, *11*, 46678−46690.
(28) He, C. B.; Cao, Y. J.; Yang, Y.; et al. Fault diagnosis of rotating machinery based on the improved multidimensional normalization ResNet. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1−11.
(29) Wu, Y. Q.; Ma, X. D. A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines. *Renewable Energy* **2022**, *181*, 554−566.
(30) Tao, C.; Tao, T.; Bai, X. J.; et al. Wind turbine blade icing prediction using focal loss function and CNN-Attention-GRU algorithm. *Energies* **2023**, *16*, 5621.
(31) Geng, Z. Q.; Chen, N.; Han, Y. M.; et al. An improved intelligent early warning method based on MWSPCA and its application in complex chemical processes. The. *Can. J. Chem. Eng.* **2020**, *98* (6), 1307−1318.
(32) Tian, Y.; Yao, H.; Li, Z. Q. Plant-wide process monitoring using weighted copula-correlation based multiblock principal component analysis approach and online-horizon Bayesian method. *ISA Transactions* **2020**, *96*, 24−36.
(33) Gao, X.; Hou, J. An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing* **2016**, *174*, 906−911.
(34) Huang, T.; Zhang, Q.; Tang, X.; et al. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artificial Intelligence Review* **2022**, *55*, 1289−1315.

https://doi.org/10.1021/acsomega.4c03757
ACS Omega 2025, 10, 8859−8869